Computer Modeling in
Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2024.052369

ARTICLE

AI-Based Helmet Violation Detection for Traffic Management System

Yahia Said[1,*], **Yahya Alassaf**[2], **Refka Ghodhbani**[3], **Yazan Ahmad Alsariera**[4], **Taoufik Saidani**[3], **Olfa Ben Rhaiem**[4], **Mohamad Khaled Makhdoum**[1] **and Manel Hleili**[5]

[1]Department of Electrical Engineering, College of Engineering, Northern Border University, Arar, 91431, Saudi Arabia

[2]Department of Civil Engineering, College of Engineering, Northern Border University, Arar, 91431, Saudi Arabia

[3]Faculty of Computing and Information Technology, Northern Border University, Rafha, 91911, Saudi Arabia

[4]College of Science, Northern Border University, Arar, 91431, Saudi Arabia

[5]Department of Mathematics, Faculty of Sciences of Tabuk, University of Tabuk, Tabuk, 71491, Saudi Arabia

*Corresponding Author: Yahia Said. Email: yahia.said@nbu.edu.sa

Received: 31 March 2024 Accepted: 28 June 2024 Published: 20 August 2024

ABSTRACT

Enhancing road safety globally is imperative, especially given the significant portion of traffic-related fatalities attributed to motorcycle accidents resulting from non-compliance with helmet regulations. Acknowledging the critical role of helmets in rider protection, this paper presents an innovative approach to helmet violation detection using deep learning methodologies. The primary innovation involves the adaptation of the PerspectiveNet architecture, transitioning from the original Res2Net to the more efficient EfficientNet v2 backbone, aimed at bolstering detection capabilities. Through rigorous optimization techniques and extensive experimentation utilizing the India driving dataset (IDD) for training and validation, the system demonstrates exceptional performance, achieving an impressive detection accuracy of 95.2%, surpassing existing benchmarks. Furthermore, the optimized PerspectiveNet model showcases reduced computational complexity, marking a significant stride in real-time helmet violation detection for enhanced traffic management and road safety measures.

KEYWORDS

Non-helmet use detection; traffic violation; safety; deep learning; optimized PerspectiveNet

1 Introduction

Traffic violation systems are being developed as the number of vehicles on the road is growing rapidly. Due to the lack of advanced detection, this increase in numbers makes it difficult for trigger-based traffic violation detection systems to keep up with the high volume of traffic. Additionally, they are not designed to detect multiple traffic violations simultaneously for a given vehicle, which leads to an increase in traffic rule violations. Traffic law violations cause a variety of traffic accidents, and traffic management in most modern urban settings presents several fundamental challenges.

The exponential growth in the number of vehicles on the roads has underscored the urgency of developing efficient traffic violation systems to ensure road safety. However, the surge in vehicle numbers poses challenges for trigger-based traffic violation detection systems, which struggle to keep pace with the high traffic volumes. Moreover, these systems often lack the capability to detect multiple violations simultaneously, leading to increased instances of traffic rule violations. Given that traffic law infractions contribute significantly to traffic accidents, particularly concerning motorcycle riders, there is an acute need for effective measures to enforce helmet usage regulations and improve road safety. Motivated by the mention, this paper aims to implement an advanced helmet violation detection system using deep learning techniques to enhance monitoring and enforcement efforts. The primary objective is to develop a highly accurate and efficient detection system capable of reducing incidents involving helmet violations, thereby lowering human and financial costs and ultimately fostering safer roads for all.

Driver distraction [1], a distraction from activities occurring within the car [2], and fatigue detection [3] present the most frequent causes of traffic accidents. Drivers aren't always conscious of their distractions or how in-car chores affect their ability to drive [4]. In keeping with this, the majority of traffic infractions, such as speeding and disobeying stop signs, are unintentional and result from a driver's inability to pay attention, rather than from a conscious choice to break the law. In addition to differences in performance level, these accidental errors could also happen as a result of disregarding traffic laws. Driving assistance systems are therefore seen as a key to preventing accidents by notifying drivers about their careless actions on the road and reminding them to be more careful.

1.35 million people worldwide die on roadways each year, based on data from the Centers for Disease Control and Prevention (CDC) [5]. Around 3700 individuals worldwide pass away every day as a result of crashes involving cars, buses, motorcycles, bicycles, trucks, or pedestrians. Over fifty percent of all fatalities involve motorcyclists, pedestrians, and cyclists. An emerging study area that is generating a lot of interest in systems for traffic control is handling traffic and traffic violation detection. Detecting traffic violations state on the road is one of the most crucial ways to relieve traffic congestion and save road users' lives. Efficient detection of on-road traffic violations is extremely important for traffic control. Traffic flow around the world is negatively impacted by traffic violations as it presents one of the primary mortality causes around the world. As a result, effective traffic violation detection systems are essential for returning traffic to normal on the roadways and protecting lives.

The world's excessive vehicle and motorbike population, growing commuter population, ineffective traffic signal management, and rider attitude make traffic infraction monitoring and control a difficult task and a significant issue. For such huge traffic volumes and traffic offenses, physical police-based traffic monitoring is insufficient. As a result, many infringers have gone undetected. As a result, the offenders cause more serious auto accidents that put both their lives and those of others in jeopardy. In order to identify and detain violators, Artificial Intelligence (AI)-based technologies must be used in place of manual involvement. Building such a precise system for tracking traffic offenses is crucial for upholding the law and keeping track of offenders. The majority of today's outdated urban traffic control systems are still manually supervised. Both heavy traffic congestion and human mistakes result from this.

The World Health Organization (WHO) found that wearing a helmet correctly can lower the possibility of fatal injuries by 42% and the risk of brain injuries by 69% [6]. Furthermore, drivers who use their phones while traveling become four times more likely to be involved in a collision than those who don't, according to data from the WHO.

The volume and type of cars are too many for traffic police monitoring to handle alone. Additionally, a lot of current systems aren't flexible enough to recognize, examine, and track the wide range of vehicle kinds, license types, dynamic traffic patterns, and street layouts [7]. Numerous cities still use antiquated traffic control systems that can't be scaled up effectively to accommodate the volume and variety of traffic. Modern technologies must be developed to monitor traffic and automate enforcement to handle these different issues allowing for more intelligent traffic control systems.

The development and improvements of sensor and Artificial Intelligence technology have led to quick advancements in the field of driving safety. In order to decrease traffic accidents, active safety technologies like anti-lock braking systems and adaptive cruise control have been widely implemented in automobiles [8].

Various traffic violations can occur while driving such as traffic light violations, traffic law violations, stop sign violations, and high speed. All these traffic violations will cause a greater number of accidents and deaths around the world. When a car approaches a junction after the traffic signal has gone red, traffic infractions occur. Although most drivers abide by traffic signs, violations can occur because of driver distraction, aggressive driving, or a conscious choice to ignore the sign. Traffic light violations are fairly common. Various accidents occur when drivers run a red light, ignore the other traffic control, and do not obey the traffic rules and low.

Road intersections are places where automobiles and pedestrians may come into conflict, raising the possibility of collisions. Even though they only make up a minor percentage of the highway system, junctions account for a sizable portion of crashes [9].

Recently, deep learning-based architectures have gained more attention as they were successfully applied to solve various computer vision tasks including indoor object detection [10], pedestrian detection [11], traffic sign detection [12], and indoor wayfinding [13].

More cars and motorcycles are being purchased, especially in urban areas, as a result of the expanding population and people's desire for comfort. Traffic may become backed up as a result, showing that traffic violations are increasing riskier everywhere. As a result, there are more accidents, which may lead to the loss of many lives, and people's awareness declines. Due to these circumstances, it is essential to create systems for detecting traffic violations in order to automate the enforcement of traffic laws and eradicate public ignorance.

Strict adherence to the laws and ongoing traffic monitoring is required to reduce the accident rate and traffic levels. Developing a new traffic rule violation monitoring system minimizes human work while ensuring that the laws are strictly respected.

The main aim of this work is to develop a helmet violation detection system that can accurately detect whether the motorcycle rider wears a helmet or not. The proposed system was developed on top of a modified version of the PerspectiveNet network [14]. The developed system will present a new tool that can ensure more respect for traffic rules and will contribute to reducing the number of fatal accidents.

The remainder of this paper is as follows: Section 2 will provide a review of the helmet violation detection works. Section 3 will detail the proposed architecture used for the helmet violation detection system. Section 4 provides all the experiments conducted in the proposed work and Section 5 concludes the paper.

2 Related Works

The number of accidents is increasing day by day due to the non-respect of traffic rules. Traffic accidents present one of the primary causes of human deaths around the world. To deal with this issue various works have been elaborated to build new systems used to maintain and follow the traffic rules.

The number of traffic fatalities is relatively high, particularly in low- and middle-income countries. The failure to wear a motorcycle helmet is one of the leading causes of traffic fatalities. Increased compliance may be aided by active law enforcement, while ubiquitous enforcement requires more police personnel and could result in traffic congestion and safety concerns.

To address the problems of traffic violations in general and the non-use of helmets by motorcycle riders' various works have been proposed in the literature.

Charran et al. [15] proposed a system used to automatically identify two-wheeler infractions for Indian road conditions, such as not wearing a helmet, using a phone while riding, triple riding, wheeling, and illegal parking, and ultimately automating the ticketing process by logging the infractions and associated vehicle number in a database. To build such a system, authors used a combination of YOLOv4 [16] and Deepsort and they achieved a mean average precision (mAP) of 98.09%.

The YOLO [17] algorithm was used in [18] to perform a system that can identify car license plates for motorcycle riders who break traffic laws. This detector will be situated at an intersection with traffic lights. The developed system uses two primary video processing methods: scanning license plates and detecting helmets and rearview glasses. Furthermore, the system will emit a warning to that particular offender. The three essential parts of the system are a camera, a computer, and a speaker. The obtained testing results from the system show that the YOLO darknet system can detect all categories with an accuracy of 93%.

One of the main reasons for fatalities among people is traffic accidents. Motorcycle accidents are among the many types of traffic accidents and can result in serious casualties. The primary form of protection for a motorcyclist is a helmet. The majority of nations mandate that motorcycle riders wear helmets, but for a variety of reasons, many people disregard the rule. A helmet and no-helmet classification system was proposed in [19]. This system was developed using deep learning techniques. An accuracy of 90% has been achieved by this system.

By wearing a helmet motorcycle riders, the number of deaths can be increased to more than 42%. Additionally, following helmet use is unresected, especially in developed countries. To ensure more traffic security, a helmet detection system was proposed [20]. This system was developed on the top of YOLOv2 network [21].

In [22], the authors demonstrated the effectiveness of computer vision and machine learning techniques that can improve helmet compliance through the automated detection of helmet violations. The system includes monitoring, classification, and biker detection to identify riders and passengers who are not wearing helmets. The system's architecture consists of a single GPU server and numerous computational clients working together to complete the task while communicating over HTTP. An accuracy of 97% has been achieved as a detection rate of helmet use.

In [23], authors proposed a method for real-time surveillance videos' automatic helmet-less bike rider recognition. The proposed method initially uses backdrop removal and object segmentation to identify bike riders in surveillance video. The binary classifier and visual cues are then used to determine whether or not the bike rider is wearing a helmet. The authors also give a consolidation method for reporting violations, which helps to increase the validity of the method. Authors have presented a performance comparison of three frequently used feature representations for classification, including the Histogram of Oriented Gradients (HOG), the Scale-Invariant Feature Transform (SIFT), and Local Binary Patterns (LBP), in order to assess their approach. According to the experimental findings, 93.80% of the real-world surveillance data have been detected.

The detection of traffic rule offenders is highly desirable, but this task is difficult because of many challenges such as occlusion, illumination, poor quality of surveillance video, fluctuating weather conditions, etc. In [24], the authors provided a system based on a deep convolutional neural network used for the automatic detection of motorcycle riders operating their vehicles without protective helmets. This system shows a detection accuracy of 92.87%.

A new system employing the use of CCTV cameras to enforce helmet use was proposed in [25]. The created application seeks to assist police in enforcing the law, ultimately changing risky behaviors and lowering the accident frequency and severity. The validation findings show that the algorithm detects the motorcycle rider without a helmet with 74% of accuracy.

To ensure safer traffic conditions, Tonge et al. [26] proposed a new traffic violation detection system that can recognize non-helmet use. The proposed work was built on top of the YOLO network. Upon the detection of violations, the relevant violators' vehicle numbers are retrieved via Optical Character Recognition (OCR), and the violators will be notified.

An Intelligent Transportation System (ITS) presents an application that contains different transportation modes and traffic control systems. ITS functions include requesting emergency assistance and using roadside equipment to enforce traffic regulations. It has been noted that many fatal motorbike accidents include people who were not wearing helmets. A challenging ITS application is the automatic helmet violation detection of motorcyclists from real-time videos. It makes it possible to identify and find bikers who do not wear helmets. Waris et al. [27] proposed a Convolutional Neural Network (CNN)-based non-helmet use detection system that can contribute to better traffic conditions and decrease the number of accidents. According to the experimental results, this system achieves an accuracy of 97.69%.

In [28], authors presented an improved disentanglement module to handle feature misalignment problems in CNN-based object detectors. These misalignments result from the network's classification and regression tasks. This technique efficiently untangles features to enhance alignment by operating in the feature pyramid network (FPN), which is the neck of the architecture. To further reduce inconsistent results and stifle subpar predictions, a response alignment technique is presented. Extensive experiments employing different backbone topologies on the two benchmark datasets MS COCO and PASCAL VOC empower the efficacy of this technique. The outcomes show notable gains in performance over the current approaches.

Various works have been proposed in the literature to reduce the number of helmet-use law-breaking, but few of them take into account to ensure a better contribution between the detection accuracy and the processing time.

In this paper, we will propose to build a new helmet and non-helmet use detection system using deep learning advantages that operate in real-time conditions and achieve better accuracy that outperforms the state-of-the-art results. The proposed system can be included in an intelligent traffic management system to ensure more respect for traffic rules and reduce the number of fatal deaths due to the non-use of helmet by motorcyclists.

3 Proposed Architecture for Helmet Violation Detection

In developing countries, motorcycles remain the most popular means of transportation, but because they are "vulnerable road users" on the road, they are more likely to be involved in collisions and result in injuries or fatalities. Bicyclists, pedestrians, and riders made up half of all fatalities on the world's roads in 2018, according to the World Health Organization's worldwide road safety status report. The main aim of this work is to build an efficient system used to detect non-helmet use. This system will maintain a high level of traffic rules respect and consideration. To this end, a helmet use detection system is developed in this paper based on a modified version of PerspectiveNet [14]. In the following, we will detail the proposed architecture used for helmet violation detection.

In order to contribute to such a system and to ensure an embedded implementation, EfficientNet v2 [29] has been used as a network backbone. EfficientNet v2 presents an optimized version of the EfficientNet family [30]. Compound coefficient scaling presents the scaling method presented by the EfficientNet architecture. The first EfficientNet architecture is shown in Fig. 1.

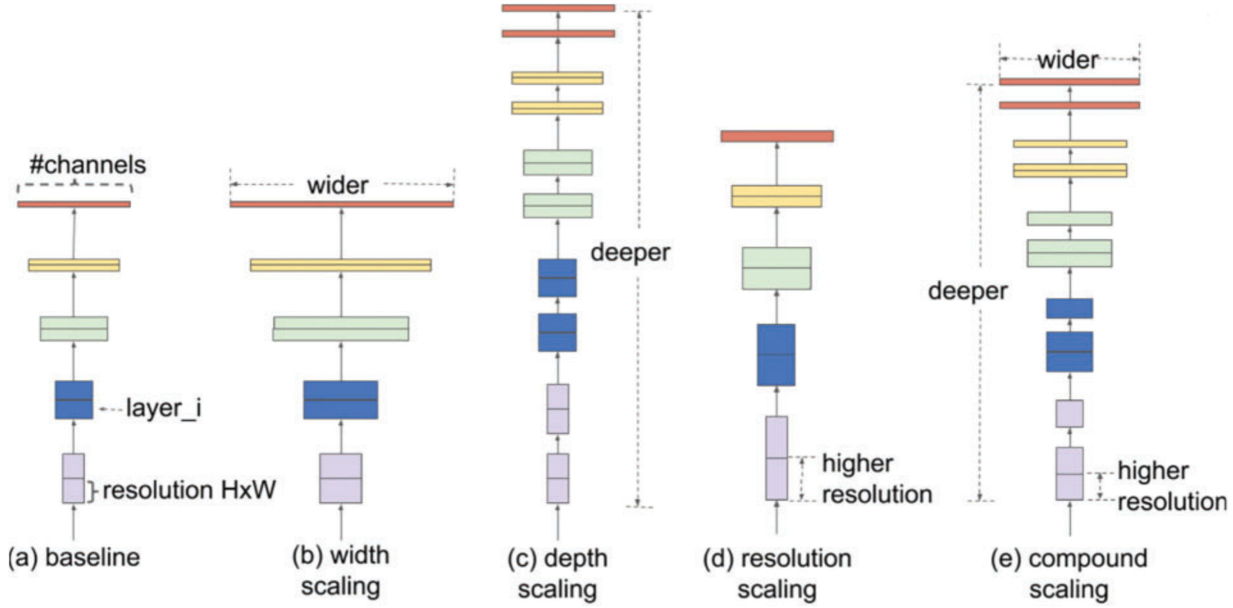

Figure 1: Compound scaling: (a): baseline, (b): width scaling, (c): depth scaling, (d): resolution scaling, (e): compound scaling

The scaling-up method is proposed in several dimensions. The EfficientNet v1 compound scaling is as follows:

- Depth: 1.20
- Width: 1.10
- Resolution: 1.15

Depthwise convolution is another strength of the EfficientNet architecture. Compared to regular convolution, this type of convolution has fewer parameters and a simpler computation. Modern accelerators do not support this type of convolution. The use of fused-MBconv is proposed in EfficientNet v2 as a solution to this issue. This type of convolution replaces the depthwise 3 × 3 and expansion conv 1 × 1 by a regular 3 × 3 convolution as presented in Fig. 2.

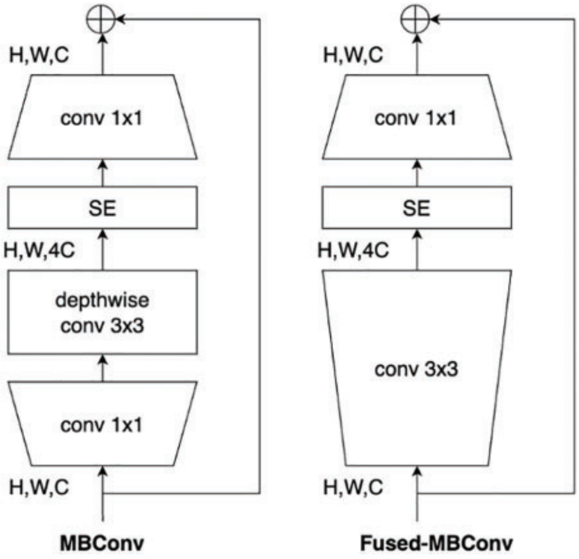

Figure 2: Difference between MBconv and fused MBconv

EfficientNet v2 architecture presents a combination of MBconv and fused MBconv in the first layers of the neural network. The second version of EfficientNet uses small expansion ratios for MBconv layers as they depend on less memory access. While the 3 × 3 convolution kernel size is used to reduce the network computation complexity. Fused-MBconv is utilized by the EfficientNet v2 design in the first phases of networks from 1 to 3. Using this fact, the process's training speed is increased while the number of floating points per second (FLOPs) and parameters are raised when the fused-MBconv is applied to the process' seventh step. Finding the ideal combination of the two MBconv and fused MBconv blocks is crucial for solving all of these issues. The EfficientNet v2 architecture uses training training-aware Neural Architecture Search (NAS) framework to search in the NAS search space for the optimal combination. The optimal network layer count, convolution kernel size, and the number of convolution blocks (MBconv and fused MBconv) are (3 × 3, 5 × 5). The "training-aware" training technique, which is used in the EfficientNet v2 architecture, tries to minimize the training procedures by pooling skip connections. In Table 1, the architecture of EfficientNet v2 is detailed.

Table 1: EfficientNet v2 architecture

Operator	Stride	Layers
Conv 3 × 3	2	1
Fused MBconv1, K3 × 3	1	2
Fused MBconv4, K3 × 3	2	4
Fused MBconv4, K3 × 3	2	4
MBconv4, K3 × 3, SE 0.25	2	6
MBconv6, K3 × 3, SE 0.25	1	9
MBconv6, K3 × 3, SE 0.25	2	15
Conv 1 × 1, & pooling & FC	–	1

The effectiveness and speed of training are significantly influenced by image size. Progressive learning combined with adaptive regularization is used in EfficientNet v2 training. The neural network is developed using poor regularization and reduced image sizes in the early training epochs. Following this procedure, the image size is increased with stronger regularization.

In order to build the proposed system used for helmet violation detection, an improved version based on a modified version of PerspectiveNet [14] has been proposed. As mentioned above the feature extraction part has been performed on the top of EfficientNet v2 [29] network while the detection part has been performed on the top of PerspectiveNet [28] network. This neural network provides an image complexity-adaptive convolutional neural network. It presents a flexible architecture according to image density and complexity. The proposed overall architecture based on a modified version of persectiveNet is presented in Fig. 3.

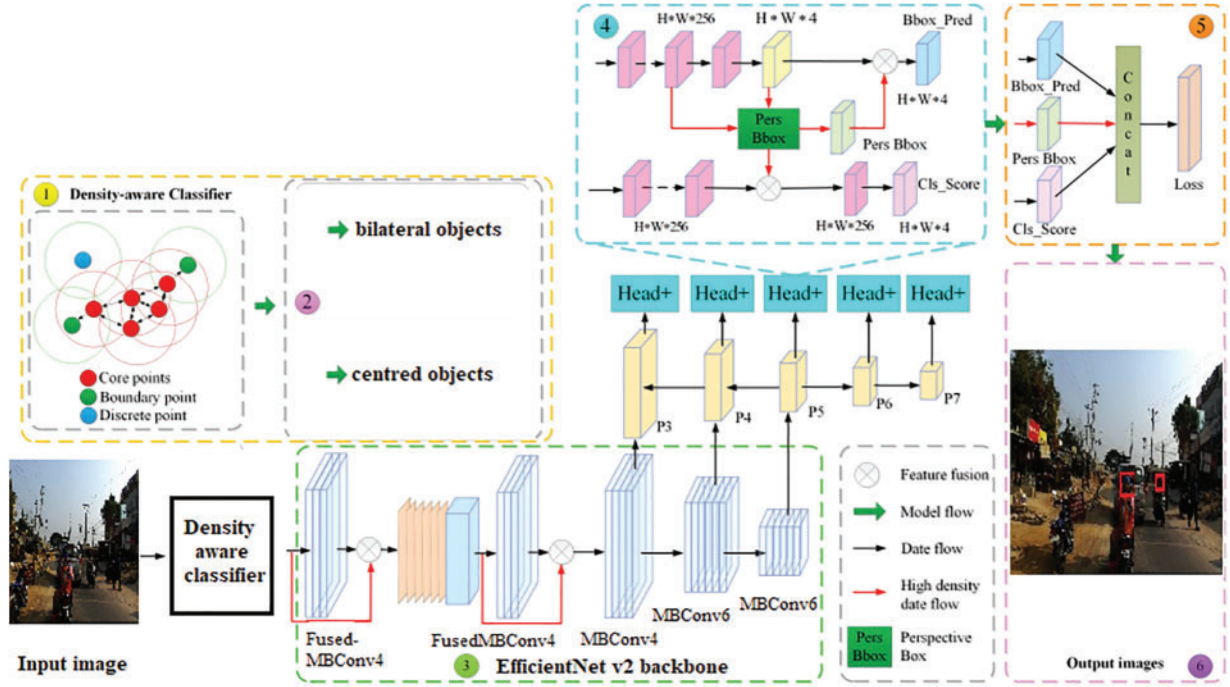

Figure 3: Proposed architecture used for helmet violation based on a modified version of PerspectiveNet [14]

A high-density network is chosen for images that have concentrated and obscured objects since it can increase the accuracy of the network. An image containing straightforward and important objects will be chosen for a low-density network. For a high-density network, a perspective box has been added to the detection head to reduce miss-detection brought on by occluded objects, and the box used for prediction will be fused to the transparent box to increase detection accuracy.

The loss function is another asset this network offers. Based on a repulsion loss of smooth function, this loss function combines characteristics loss and classification loss in contributing to the neural network's total loss function. A new neural network architecture called PerspectiveNet [14] offers a density-aware method based on variofocalNet. The six main components that make up the PerspectiveNet network are listed below.

Density-aware classifier: In particular, this section is utilized to calculate the object's density, and a convolutional neural network with adaptive learning will be developed. For input photos featuring important items, a low-density network will be used, and a high-density network will be chosen to improve the model's accuracy.

Data classification part: During the training process, the data will be divided into two main categories: bilateral objects and central objects and according to the image complexity degree with objects, an adaptive convolutional neural network will be designed. For bilateral occluded objects, a high-density network will be selected in order to improve the network's accuracy. For central images, a low-density network will be used.

Adaptive neural network: In traffic environments, different objects can interfere which can cause high-level feature loss. After the features extraction step via EfficientNet v2 backbone in the proposed work, the extracted features will be fed into the feature pyramid network FPN [31] to improve the detection accuracy. The FPN network provides a bottom-up structure with different layers (C1, C2, C3, C4, C5). The network is able to do detection at various levels and scales thanks to these many layers.

For a high-density network, the information from the C1 and C2 levels is fused before being passed through to the C3 layer with the density information to ensure that shallower feature information from deeper layers. As a result, the bottom-up structure is used to combine various informational levels. C3 and C4 layers will be fused via a 2× up operation as shown in Fig. 4. C1 and C2 will be fused with C3 via a 1 × 1 and 3 × 3 convolutions to contribute for (P1, P2, P3, P4 and P5) layers. The PerspectiveNet head consists of regression and classification parts. The features map of each FPN layer is fed into the local subnet. The characteristic map of 256 channels is produced using three 3 × 3 convolution layers with Rectified Linear Unit (RELU) as an activation function. The feature map is then convolved once more to produce a distance vector (l′, t′, r′, b′) which is used to indicate the initial bounding box.

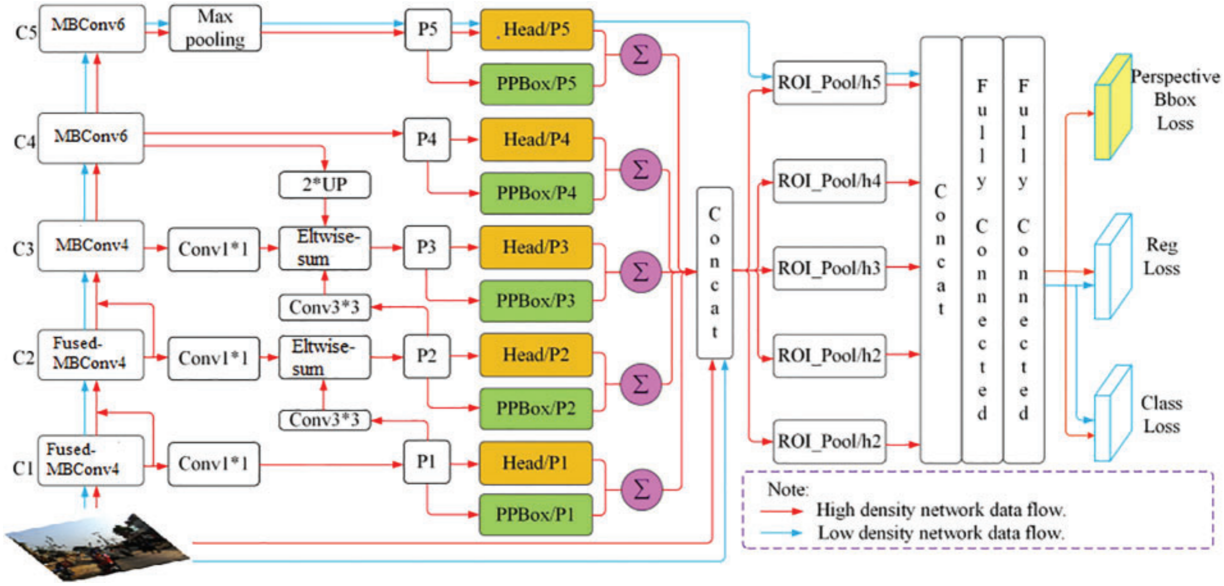

Figure 4: The adaptive neural network based on density perception. The backbone is created on top of EfficientNet V2. The image presents two paths: the high-density network (red) or the low-density network (blue) [14]

In the other branch, the perspective idea is used to determine the center coordinates (x′, y′) of the PPers Bbox circle, as well as the length $d_s^{n'}$ and width $r_s^{n'}$distances from the center to the box, in order to obtain the distance vector (x′ + $d_s^{n'}$, y′ + $r_s^{n'}$, x′ + $d_s^{n'}$, y′ + $r_s^{n'}$) of the perspective box.

The starting bounding box and visible box are merged in accordance with the distance transformation ratio (l, t, r, b) to provide the refined bounding box. The classification branch forecasts the Intersection over Union (IoU) Aware Classification Scores (IACS). It is organized similarly to the localization subnet, with the exception that each spatial location creates a vector of categories, each of which contains a representation of the object's placement confidence and existence confidence. For each of the levels of the feature layer (P1, P2, P3, P4, P5) of the FPN, we apply IACS prediction for the class of position in order to improve detection accuracy. Then, we mix the position prediction with the classification prediction to obtain the object's location and classification.

Perspective box: To improve item recognition and placement accuracy as well as to solve the issue of missed detection brought on by occlusion, the perspective box has been combined with the detection head. Two rays are drawn using this method, each one being 1/3 of the dimension of the input image. The ray equation is r = $\frac{rs}{f}$d + r_s and r = $\frac{rs}{ds}$d, where r_s is the distance between the center of the circle and the short side of GroudTruth, ds is the distance between the circle center and the big side of GroudTruth, and f is the focal length. The adjacent perspective box's length and width can be calculated as Eq. (1).

$$\begin{cases} d_s' = \dfrac{f \cdot d_s}{d_s - f} \\ r_s' = \dfrac{f \cdot r_s}{r_s - f} \end{cases} \tag{1}$$

where d_s' is half the projected length of the perspective and r_s' is half the predicted width of the perspective. The recursive technique, as demonstrated in Eqs. (2) and (3), may be used to calculate the size of all perspective boxes.

$$\begin{cases} d_s^{(n-1)'} & \textit{if n is even} \\ d_s' = \dfrac{f \cdot d_s^{(n-1)'}}{d_s^{(n-1)'} - f} & \textit{if n is odd} \end{cases} \tag{2}$$

$$\begin{cases} r_s^{(n-1)'} & \textit{if n is even} \\ r_s' = \dfrac{f \cdot r_s^{(n-1)'}}{r_s^{(n-1)'} - f} & \textit{if n is odd} \end{cases} \tag{3}$$

where d (n−1)′s and r (n−1)′s present the length and width of n−1 perspective boxes.

Loss function: As stated in Eq. (4), the overall loss of the perspective model is made up of four parts.

$$loss = \alpha L_{GT} + \beta L_{PPloss} + L_{Bbox} + L_{CLS} \tag{4}$$

where L_{GT} stands for Ground Truth loss, L_{PPloss} stands for Perspective loss, l_{Bbox} stands for Bounding Box loss, and L_{cls} stands for Classified loss. α is a modifying factor and β a focusing parameter.

Due to how closely spaced apart the occluded objects are, during the NMS (non-maximum suppression) process, they were easily cleared, leading to missed detection. PerspectiveNet is modeled after [27] in order to handle this issue, where the perspective loss consists of L_{GT} and L_{PPloss} exclusion terms, which need a predicted box to exclude other nearby ground truth objects as well as other predicted boxes with other designated items. As weights to account for auxiliary losses, the coefficients α and β are used. It is possible to determine perspective loss L_{PPloss} using Eq. (5). A smooth ln operation, which can be derived using Eqs. (6) and (7), smooths the perspective loss.

$$l_{PPloss} = \frac{\sum_{i=1}^{n}(\min(\sqrt{|x_{gt}^{i} - x_{pp}^{i}|^2 + |y_{gt}^{i} - y_{pp}^{i}|^2}))}{\sum_{i=1}^{n}(\max(\sqrt{|x_{gt}^{i} - x_{pp}^{i}|^2 + |y_{gt}^{i} - y_{pp}^{i}|^2}))} \tag{5}$$

$$l_{PPloss} = \frac{\sum_{i \neq j}(smooth_{ln}(l_{PPloss}))}{\sum_{i \neq j}(l_{PPloss})} \tag{6}$$

$$smooth_{ln} = \begin{cases} -\ln(1-x), & x \leq \sigma \\ \dfrac{x-\sigma}{1-\sigma} - \ln(1-\sigma), & x > \sigma \end{cases} \tag{7}$$

where (x_{gt}, y_{gt}) is the center coordinate of the GroudTruth. The perspective box center coordinate is (x_{pp}, y_{pp}), and i is the original input image index. $\sigma \in to$ [0, 1] is the smooth parameter used to alter the sensitivity of the repulsion loss to outliers.

As in Varifocal Network (VFNet) [32], PerspectiveNet employs classification and bounding box losses, and the overall loss of the perspective net model is represented in Eq. (8).

$$\begin{aligned} \mathrm{loss} = {} & \frac{1}{\mathrm{N_{pos}}}\sum_{\mathrm{i}} \mathrm{L_{cls}} + \frac{1}{\mathrm{N_{pos}}}\sum_{\mathrm{i}}\sum_{\mathrm{c}}\left(\alpha \mathrm{L_{gt}} + \beta \mathrm{L_{PPloss}}\right) + \frac{\lambda 0}{\mathrm{N_{pos}}}\sum_{\mathrm{i}} \mathrm{q_c^*, iL_{Bbox}}\left(\mathrm{Bbox_i', Bbox_i^*}\right) \\ & + \frac{\lambda 1}{\mathrm{N_{pos}}}\sum_{\mathrm{i}} \mathrm{q_c^*, iL_{Bbox}(Bbox_i, Bbox_i^*)} \end{aligned} \tag{8}$$

where L_{Bbox} presents the distance-IoU loss (DIoU), $Bbox_i$, $Bbox_i'$ and $Bbox_i^*$ respectively presents the initial, refinement, and ground truth bounding boxes. λ0 and λ1 refer to the balancing parameters in the loss function. The number of categories utilized to normalize total losses is denoted by $\mathrm{N_{pos}}$.

Objects detection diagram: As a network output, the different class objects that were defined during the training step will be detected and each object will be defined by its bounding box.

The activation function makes a significant contribution to improving neural network design performance and outcomes. The scaled polynomial constant unit activation function (SPOCU) [33] was developed to overcome several challenges and minimize the computational complexity of deep learning models. Self-Paced Output Control Unit (SPOCU) outperformed state-of-the-art activation functions such as Scaled Exponential Linear Unit (SELU) [34] and RELU [35]. In Eq. (9), the SPOCU activation function is employed as the activation function for the EfficientNet v2 backbone.

$$S(x) = \alpha h\left(\frac{x}{\gamma} + \beta\right) - \alpha h(\beta) \tag{9}$$

where $\beta \epsilon (0, 1), \alpha, \gamma > 0$ and

$$\mathrm{h(x)} = \begin{cases} r(c), & x \geq c \\ r(x), & x \epsilon [0, c] \\ 0, & x < 0 \end{cases}$$

with $r(x) = x^3(x^5 - 2x^4 + 2)$ and $1 \leq c < \infty$.

4 Experiments and Results

4.1 Experiments Details

In this section, we will go through all of the details and the experiments conducted to contribute to this work. In order to build the proposed helmet violation detection system, the India driving dataset (IDD) [36] has been used.

The starting learning rate is 0.0001. The final learning rate is 5e-04. 30 training epochs are used and each training epoch contains 8000 iterations, and the IoU threshold is 0.5. In this work, the $\lambda 0$ and $\lambda 1$ in Eq. (8) are set to 1.5 and 2.0, respectively. In order to contribute to better detection results various experiments have been performed. We changed the training batch size as well as the network optimizer to increase the system accuracy. For this, we used two famous network optimizers: Adaptive Moment Estimation (Adam) and Root Mean Square Propagation (RMSProp). Training batch sizes have been set into 6 and 8. Table 2 provides all the experiment settings used in this work.

Table 2: Experiments settings

Training iterations	8000
Number of epochs	30
Learning rate	0.0001
Training set	65%
Testing set	25%
Validation set	25%
Train batch size	8/6
Loss function	Pploss
Network optimizer	RMSProp/ADAM
Activation function	RELU/SPOCU

4.2 Data Preparation Details

Training, validation, and testing experiments have been conducted using the IDD as mentioned above. This dataset provides two main parts: one for segmentation issues and a second for detection problems. In the proposed work we focused our attention on the detection dataset. The images in the dataset were captured by a front-facing camera mounted on an automobile. It consists of 46,588 images divided into train, validation, and test parts. The IDD detection dataset provides 15 labels that are relevant to driving scenarios. The main aim of the proposed work is to build an efficient system used to detect helmet and non-helmet use and to decrease the number of this violation that presents one of the most frequent mortality causes in the world. To fill this end, in the proposed work, we used a subset from the IDD dataset (5000 images) and we performed our labeling which consists of two main classes (helmet and no helmet). The image labeling has been manually performed using the labelImg tool. For the purposes of this work, we selected a subset of 5000 images from the IDD dataset and focused on annotating them into two primary categories: “helmet” and “no helmet.” This decision was driven by the specific goal of enhancing safety measures in traffic environments by detecting helmet usage among motorcyclists and riders. By concentrating on a smaller, manageable subset of the dataset, we were able to ensure thorough and accurate annotations, which are crucial for training a reliable detection model. The choice to limit the classification to two categories simplifies the problem, allowing for more precise model training and evaluation, which is essential for developing an effective and efficient system capable of identifying helmet compliance in real-time traffic scenarios. Various excessive experiments have been conducted in order to obtain better detection results and contribute to better accuracies. In the proposed experiments, we used the following training protocol: 50% for training, 25% for validation, and 25% for testing.

The most frequent problem in deep learning-based networks is a class imbalance which leads to different problems including bad detection results. To address this problem, a thorough examination of the factors contributing to misidentified outcomes in detecting helmet and no helmet scenarios is crucial for improving the accuracy and reliability of the system. Factors such as the angle of photography can significantly affect the visibility and distinguishability of helmets, as certain angles may obscure the helmet or create visual distortions. Lighting conditions also play a pivotal role, with poor lighting or excessive glare potentially leading to incorrect classifications. Additionally, obstructions by surrounding objects, like other vehicles, pedestrians, or environmental elements, can partially or completely hide the helmet, leading to false negatives. To address these problems, various data augmentation techniques have been applied including horizontal flipping, vertical flipping, image translation, random cropping, brightness adjusting, and random translation. The IDD dataset used during the proposed experiments presents very challenging data as it offers real-world conditions of traffic environments. Also, it is very useful for helmet traffic violations.

4.3 Evaluation Metrics

In order to ensure an in-depth study of the proposed work, different evaluation metrics have been adopted. We tested our model's performance using a variety of measures, including F1-score, recall, precision, and accuracy. All these evaluation metrics can be calculated as Eqs. (10)–(13).

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Accuracy = \frac{TP + TN}{TN + FN + FP + TP} \tag{13}$$

where the number of correctly predicted positive samples is denoted by TP (True Positive), The number of correctly predicted negative samples is denoted by TN (True Negative), the number of wrongly anticipated negative samples is denoted by FP (False Positive), and the number of mistakenly predicted positive samples is denoted by FN (False Negative).

4.4 Detection Results

In order to study the effectiveness of the proposed helmet violation detection system, during the training process, we changed the network optimizer. For this fact, we used two well-known optimizers: Adam and RMSProp. Table 3 provides the obtained performances for the two optimizers used.

As presented in Table 3, by using Adam as a network optimizer, the obtained results have been improved. As an example, the accuracy was increased by around 2.5%.

Table 3: Obtained testing accuracies for different optimizers

Optimizer	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
RMSProp	92.7	90.3	91.2	90.7
Adam	95.2	92.1	93.3	92.6

To increase the effectiveness of the suggested helmet violation detection system, we modified the batch size and tested its performance. Batch size is one of the key hyperparameters for deep learning models. As a result, the accuracy of testing is tested for various batch sizes. The accuracy of the tests is shown in the table below for two different batch sizes, 6 and 8. The findings in Table 4 indicate that increasing the batch size from 6 to 8 resulted in the highest testing accuracy.

Table 4: Testing accuracies achieved for different batch sizes

Batch size	6	8
Accuracy (%)	93.2	95.2

4.5 Ablation Study

To improve detection results and to ensure better traffic conditions by developing an efficient system used for helmet violation detection, the PerspectiveNet backbone employed in this study is the EfficientNet v2. The network backbone was trained with two alternative activation functions: RELU and SPOCU [33]. Table 5 shows the detection results obtained when the network activation function was changed and Adam was used as a network optimizer.

Table 5: Activation function's impact on detection performance

Activation function	Accuracy (%)
RELU	93.5
SPOCU	95.2

The network detection performance has been enhanced by roughly 2% by switching the network activation function from RELU to SPOCU. SPOCU makes it possible for neural network topologies to enhance detection outcomes.

PerspectiveNet network was trained and tested using two distinct backbones, Res2Net 101 [37] and EfficientNet v2, to get better performance with the least amount of computation complexity. Table 6 shows the results achieved.

The usage of EfficientNet v2 as a network optimizer in the proposed study significantly enhanced the results. As seen in Table 6, the EfficientNet v2 backbone contributes to better detection precision and less computation complexity. Based on the obtained results, the developed helmet violation detection system can be implemented on low-end devices. It reduced the number of parameters and FLOPs utilized by the PerspectiveNet design significantly.

Table 6: Backbone modification impact on neural network computation complexity and detection performances

Backbone	FLOPS (B)	Parameters (M)	Accuracy (%)
Res2Net 101 [37]	191.83	148.87	92.9
EfficientNet v2 [29]	12.53	32.43	95.2

5 Conclusion

Road accidents are among the leading causes of human death. As a result of human irresponsibility, the frequency of traffic accidents continues to rise worldwide. Several individuals lose their lives in motorcycle accidents, primarily because they are not wearing helmets. Automatic helmet violation detection presents a demanding application that should be developed to address this problem and to maintain the traffic rules respected. To fill this end, In the proposed work a deep learning-based system used for helmet violation detection is proposed. The developed system was built on top of a modified version of PerspectiveNet. The efficientNet v2 backbone has been used for feature extraction. The developed helmet detection system based on the proposed PerspectiveNet version has shown better performances with lower computation complexity than the original PerspectiveNet version. The proposed system can detect whether the motorcycle user uses a helmet or not in a very accurate way.

According to the experiment's achievements, the proposed helmet violation detection system provides 95.2% detection accuracy with much lower computation complexity. Despite the high detection results obtained by the proposed helmet detection system, this work presents some limitations. The modified version of PerspectiveNet can be evaluated using various types of datasets to ensure its robustness and generalization. A full study about real-world conditions deployment should be performed for practical deployment in traffic monitoring scenarios.

As a future work, the proposed helmet violation detection system developed on a modified version of PerspectiveNet will be implemented on an embedded device.

Acknowledgement: The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University for funding this work through Research Group No. (RG-NBU-2022-1234).

Funding Statement: This research was funded by the Deanship of Scientific Research at Northern Border University, Arar, Kingdom of Saudi Arabia through Research Group No. (RG-NBU-2022-1234).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Yahia Said, Yahya Alassaf, Yazan Ahmad Alsariera; data collection: Taoufik Saidani, Refka Ghodhbani, Mohamad Khaled Makhdoum; analysis and interpretation of results: Olfa Ben Rhaiem, Manel Hleili, Mohamad Khaled Makhdoum; draft manuscript preparation: Yahia Said, Yazan Ahmad Alsariera, Yahya Alassaf, Manel Hleili. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data will be made available on request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Eraqi HM, Abouelnaga Y, Saad MH, Moustafa MN. Driver distraction identification with an ensemble of convolutional neural networks. J Adv Transport. 2019 Feb 13;2019:1–12.
2. Vismaya UK, Saritha E. A review on driver distraction detection methods. In: International Conference on Communication and Signal Processing (ICCSP); 2020 Jul 28; Melmaruvathur, India; p. 483–7.
3. Ayachi R, Afif M, Said Y, Abdelali AB. Drivers fatigue detection using efficientdet in advanced driver assistance systems. In: International Multi-Conference on Systems, Signals & Devices (SSD); 2021 Mar 22; Monastir, Tunisia; p. 738–42.
4. Trivedi MM, Gandhi T, McCall J. Looking-in and looking-out of a vehicle: computer-vision-based enhanced vehicle safety. IEEE Trans Intell Transp Syst. 2007 Feb 26;8(1):108–20.
5. https://www.cdc.gov/. [Accessed 2024].
6. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries/. [Accessed 2024].
7. Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric. In: IEEE International Conference on Image Processing (ICIP); 2017 Sep 17; Beijing, China; p. 3645–9.
8. Zhu S, Fan X, Qi G, Wang P. Review of control algorithms of vehicle anti-lock braking system. Recent Pat Eng. 2023 Mar 1;17(2):30–45.
9. Choi EH. Crash factors in intersection-related crashes: an on-scene perspective; 2010 Sep. (No. HS-811 366).
10. Afif M, Ayachi R, Said Y, Pissaloux E, Atri M. Indoor object c1assification for autonomous navigation assistance based on deep CNN model. In: IEEE International Symposium on Measurements & Networking (M&N); 2019 Jul 8; Catania, Italy; p. 1–4.
11. Mounsey A, Khan A, Sharma S. Deep and transfer learning approaches for pedestrian identification and classification in autonomous vehicles. Electronics. 2021 Dec 18;10(24):3159.
12. Triki N, Karray M, Ksantini M. A real-time traffic sign recognition method using a new attention-based deep convolutional neural network for smart vehicles. Appl Sci. 2023 Apr 11;13(8):4793. doi:10.3390/app13084793.
13. Crabb R, Cheraghi SA, Coughlan JM. A lightweight approach to localization for blind and visually impaired travelers. Sensors. 2023 Mar 1;23(5):2701. doi:10.3390/s23052701.
14. Li CJ, Qu Z, Wang SY. PerspectiveNet: an object detection method with adaptive perspective box network based on density-aware. IEEE Trans Intell Transp Syst. 2023 Jan 31;24(5):5419–29. doi:10.1109/TITS.2023.3240616.
15. Charran RS, Dubey RK. Two-wheeler vehicle traffic violations detection and automated ticketing for Indian road scenario. IEEE Trans Intell Transp Syst. 2022 Jul 12;23(11):22002–7. doi:10.1109/TITS.2022.3186679.
16. Bochkovskiy A, Wang CY, Liao HY. YOLOv4: optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. 2020 Apr 23.
17. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016; Las Vegas, NV, USA; p. 779–88.
18. Ferdian R, Sari TP. Identification of motorcycle traffic violations with deep learning method. In: International Symposium on Information Technology and Digital Innovation (ISITDI); 2022 Jul 27; p. 146–9. doi:10.1109/ISITDI55734.2022.9944502.

19. Raj KD, Chairat A, Timtong V, Dailey MN, Ekpanyapong M. Helmet violation processing using deep learning. In: International Workshop on Advanced Image Technology (IWAIT); 2018 Jan 7; Chiang Mai, Thailand; p. 1–4.
20. Sridhar P, Jagadeeswari M, Sri SH, Akshaya N, Haritha J. Helmet violation detection using YOLO v2 deep learning framework. In: International Conference on Trends in Electronics and Informatics (ICOEI); 2022 Apr 28; Tirunelveli, India; p. 1207–12.
21. Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017; Honolulu, HI, USA; p. 7263–71.
22. Chairat A, Dailey M, Limsoonthrakul S, Ekpanyapong M, Dharma Raj KC. Low cost, high performance automatic motorcycle helmet violation detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision; 2020; Snowmass village, Colorado; p. 3560–8.
23. Dahiya K, Singh D, Mohan CK. Automatic detection of bike-riders without helmet using surveillance videos in real-time. In: International Joint Conference on Neural Networks (IJCNN); 2016 Jul 24; Vancouver, Canada; p. 3046–51.
24. Vishnu C, Singh D, Mohan CK, Babu S. Detection of motorcyclists without helmet in videos using convolutional neural network. In: International Joint Conference on Neural Networks (IJCNN); 2017 May 14; Anchorage, Alaska; p. 3036–41.
25. Wonghabut P, Kumphong J, Satiennam T, Ung-Arunyawee R, Leelapatra W. Automatic helmet-wearing detection for law enforcement using CCTV cameras. InIOP Conf Series: Earth and Environ Sci. 2018 Apr 1;143(1):012063.
26. Tonge A, Chandak S, Khiste R, Khan U, Bewoor LA. Traffic rules violation detection using deep learning. In: International Conference on Electronics, Communication and Aerospace Technology (ICECA); 2020 Nov 5; Coimbatore, India; p. 1250–7.
27. Waris T, Asif M, Ahmad MB, Mahmood T, Zafar S, Shah M, et al. CNN-based automatic helmet violation detection of motorcyclists for an intelligent transportation system. Math Probl Eng. 2022 Oct;(1):8246776.
28. Lin W, Chu J, Leng L, Miao J, Wang L. Feature disentanglement in one-stage object detection. Pattern Recognit. 2024;145:109878.
29. Tan M, Le Q. EfficientNet v2: smaller models and faster training. In: International Conference on Machine Learning; 2021 Jul 1; p. 10096–106.
30. Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning; 2019 May 24; Long Beach, CA, USA; p. 6105–14.
31. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017; Honolulu, HI, USA; p. 2117–25.
32. Zhang H, Wang Y, Dayoub F, Sunderhauf N. VarifocalNet: an IoU-aware dense object detector. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2021; Nashville, TN, USA; p. 8514–23.
33. Kiseľák J, Lu Y, Švihra J, Szépe P, Stehlík M. SPOCU: scaled polynomial constant unit activation function. Neural Comput Appl. 2021;33:3385–401.
34. Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Self-normalizing neural networks. Advances in Neural Information Processing Systems. 2017;30:1–10.
35. Agarap AF. Deep learning using rectified linear units (ReLU). arXiv preprint arXiv:1803.08375. 2018 Mar 22.
36. Varma G, Subramanian A, Namboodiri A, Chandraker M, Jawahar CV. IDD: a dataset for exploring problems of autonomous navigation in unconstrained environments. In: Winter Conference on Applications of Computer Vision (WACV); 2019 Jan 7; Waikoloa Village, HI, USA; p. 1743–51.
37. Gao SH, Cheng MM, Zhao K, Zhang XY, Yang MH, Torr P. Res2Net: a new multi-scale backbone architecture. IEEE Trans Pattern Anal Mach Intell. 2019 Aug 30;43(2):652–62.