SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V11I5P110

Volume 11 Issue 5, 102-118, May 2024
© 2024 Seventh Sense Research Group®

Original Article

Validation and Performance Contrast of Deep Neural Network Based Mechanism for Real-Time Automatic Safety Helmet Detection

Dasari Naga Vinod[1], N. Kapileswar[2], Judy Simon[2], Phani Kumar Polasi[3], B. Padmavathi[2], Partho Adhikari[4]

[1]*Department of Electronics and Communication Engineering,*
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Tamilnadu, India.
[2]*Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Tamil Nadu, India.*
[3]*Department of Electronics and Communication Engineering, Chalapathi Institute of Technology, Andhra Pradesh, India.*
[4]*Department of Electrical Engineering and Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan.*

[1]*Corresponding Author : dasaarinagavinod@gmail.com*

Received: 07 March 2024 Revised: 07 April 2024 Accepted: 05 May 2024 Published: 29 May 2024

Abstract - *In the fourth industrial revolution, more people were working in production sites, including traffic areas, chemical plants, nuclear reactors, and building sites, raising worries about worker safety. The Head is the most important part of the body, and there have not been many studies done on detecting helmet use across a variety of units in hazardous environments. This research constructed a revolutionary deep learning mechanism called YOLOv8.0 that integrates object detection, key point localization, and basic rule-based reasoning to solve this problem. Moreover, it presents an active helmet-wearing detection method and a dataset created from scratch for multifunctional use applications. The three research questions that guide the process are, (i) Is it possible to identify certain classes in any video?, (ii) Can the model be used to identify helmets across numerous sites, and (iii) Can detections be made in challenging environmental circumstances in real-time? first created a dataset in Yolo format and augmented the photos to produce a second, more generic dataset. Next, the proposed datasets on 5 different versions of the YOLOv8.0 pre-trained model because YOLOv8.0 is an anchor-free mechanism; it detects object's centers straightly rather than its redeem through a known anchor box. On the augmented dataset original dataset, the YOLOv8l version received the best mAP score among the other versions, scoring approximately 95% and 81% respectively.*

Keywords - *Computer vision, Image augmentation, Object detection, Safety helmet, SOTA DNN, YOLOV8.*

1. Introduction

Every year, more than 1.3 million individuals expire on the globe's roadways, while two fifty million suffer minor injuries. The Worldwide Status Survey on Highway Safety is the initial comprehensive evaluation of the state of road safety in 178 nations based on data from a standardized survey. The findings reveal that injuries sustained in crashes continue to be a dominant public health problem, especially in low as well as middle-income nations.

Pedestrians, bicycles, as well as bikers account for over half of all road fatalities, underscoring the importance of giving these road users greater prominence in road safety programs. The findings indicate that many countries' road safety legislation should be expanded, and implementation should be tightened. The conclusions of the Worldwide Status Report on Highway Safety reveal that much more work is required to make the globe's roadways safer [1].

Motorcycles have become the primary mode of transportation in some impoverished communities, owing to limitations in urban infrastructure as well as economic constraints, and the fatality rate from road traffic in these places is around 3 times that of refined locations. Motorbike traffic accident mortality led to 43% as well as 0.36 of all traffic accident deaths in Southeast Asia as well as the Western Pacific area, respectively, in India, Vietnam, Indonesia, and other nations. Figure 1 is a graphical representation of fatal workplace accidents in the United Kingdom (2016-2021). According to the WHO, motorbike head lacerations are the dominant cause of mortality.

Motorbikers who wear helmets appropriately can cut their chance of mortality by 42% and their risk of brain injury by 69%. As a result, motorcycle riders must wear helmets [2]. However, for a variety of reasons, the rate of helmet use in certain developing nations has been extremely low.

 This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

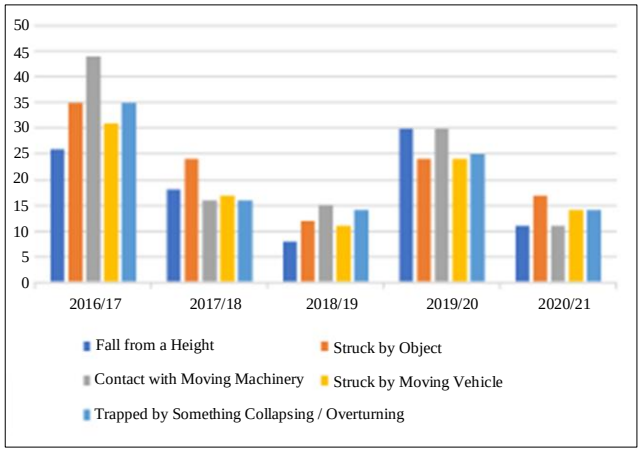


Fig. 1 Depicts the most common forms of fatal workplace accidents in the United Kingdom [7]

The Indian regime has suggested several fines under the Motor Vehicles (Amendment) Law of 2019 to boost the use of helmets. A biker who fails to wear a helmet receives a penalty of 1000 rupees and loses his or her driving license for 3 months under Section 194d. Even if these rules are obeyed, individuals would strive to avoid being arrested by traffic cops, and severe law enforcement also necessitates a large number of cops, which is both times taking as well as expensive.

To minimize the number of casualties in motorbike traffic mishaps, it is important to design a robotic helmet-detection motorcycle technology powered by deep learning. Convolutional neural networks have been widely employed in present years, with the expeditious growth of DL, like semantic disjunction as well as object identification, all of which have achieved significant advancements. CNNs are used to build semantic segmentation, a sensor-level vision job rapidly. Wang et al. [3] present a strategy for increasing classification efficiency from synthetic to real-life information that is weakly guided by adversarial domain adaptation.

Deep learning-positioned object recognition techniques are currently classified into two classes: one-stage techniques as well as two-stage techniques. Two-stage techniques are the faster R-CNN [4] as well as the mask R-CNN [5]. These techniques offer high precision for recognition, but their performance is low, making real-time detection difficult. To resolve this issue, authors presented one-stage mechanisms such You Only Look Once (YOLO) [6, 7], Single Shot Detector (SSD) [8, 9], and RetinaNet [8] positioned on anchor as well as Adaptive Training Sample Selection (ATSS) [10], as well as RepPoints [11] positioned on anchor free. The YOLO algorithm is available in eight different flavours. In this work, the fifth version of the YOLO detection method, YOLOv5, is applied [12]. It is both quick and exact, with very few model parameters. The shortest variant is only 7.3M long. To address the above issues, a new YOLOV8 was implemented.

The main components of this article are illustrated as follows:

- Created a new corpus dataset from scratch and introduced image augmentation
- Developed one of the SOTA models enabled by YOLOV8.
- Deployed the model to identify helmet-wearing in real-time by accessing the webcam.
- Compared the SHEL5K dataset (6 classes) with the proposed original and augmented dataset with 8 classes enabled by YOLOV8n, YOLOV8s, YOLOV8m, YOLOV81, and YOLOV8x.
- Associated with various versions of YOLO, the YOLOV8x mechanism achieved an average mAP of 0.9 from the original dataset and obtained 0.953 mAP from an augmented dataset.

The summary of this article is as follows. The following sections focus on recent research leveraging deep-learning algorithms to identify safety helmets. The data-collecting process, the method used for identifying safety helmets, and the different performance assessment criteria for the algorithms are all described in Section 3. Experimental outcomes from the present research are illustrated in Section 4 as well as the denouement is laid out in Section 5.

2. Literature Survey

Researchers have suggested deep learning-based approaches in a variety of areas in recent years, including medicine, agriculture, defence, and others. In [13], the background removal technique as well as the SMO predictor, are utilized to recognize motorbikes in movies. Hand-Crafted lineaments, as well as CNN, are then utilized to distinguish between helmets as well as no helmets. At last, it appears that in terms of precision, CNN exceeds manual features. To extract the moving object in the video frame, the authors of [14] apply adaptive background subtraction. Then, CNN employs this technique to classify bikers as moving particles. Moreover, they continue to enable CNN to categorize motorcycle top quarter areas to verify motorists' absence of helmets. Gaussian mixture technique is utilized [15] to segment as well as recognize foreground objects.

The system then employs a faster region-positioned CNN to identify motorbikes in the specified foreground elements, assuring the presence of riders. Subsequently, the faster R-CNN was employed to determine whether or not motorcyclists were wearing helmets. Even though the helmet recognition technique is employed in conventional background removal, it remains employed in the motorbike recognition phase to get the foreground goal, and this will be very poor in an overflowing scene. Despite motorcycle recognition is not reported, it is recommended in [16-18] to use the YOLOv3 technique to detect if the biker is wearing a helmet.

They utilized the YOLOv3 approach to identify the motorcycle as well as the individual in the visual [19, 20]. Then they calculated the region in which they overlapped the bounding box encompassing the biker as well as the human to identify who was driving the bike. In the end, the YOLOv3 method was used to identify whether the rider was equipped with a helmet or not.

Nevertheless, riders, as well as motorcycles, are highly merging in terms of surveillance of traffic, so identifying riders separately is unnecessary. In [21], the motorcycle area is detected using the SSD or YOLOv3 approach. The top half of the image is instanced as well and the categorization technique is utilized to discriminate among Helmet as well as non-helmet.

Similarly, if more than a single individual is riding on a motorcycle, the classification method is rendered worthless. They analyze the motorcycle as well as the biker as a whole in [22-24] and then utilize the CNN technique to decide if the person riding the bike is wearing a helmet. The precision of this single-step coarse-grained identification technique is relatively low.

YOLO is a simple one-stage recognition of objects approach. The identification challenge is recast as a regression issue. Rather than computing RoI, the regression method provides the organization of the bounding container and a likelihood for every class instantly. When contrasted to the faster R-CNN, it significantly enhances identification speed.

In 2020, ultralytics released YOLOv5, the fifth edition of YOLO that beats all previous versions with respect to speed as well as precision. Using the depth_multiple as well as width_multiple variables, the YOLOv5 method modifies the width as well as the depth of the backbone network, generating 4 approaches: YOLOv5m, YOLOv5s, YOLOv5x, and YOLOv5l [25].

3. Proposed Methodology

3.1. Experimental Setup

In order to construct the deep learning technique, this task entailed installing Anaconda to pre-process the data and using Google Colab to train and assess the model. The CUDA and CuDNN libraries required for GPU support utilizing PyTorch will be installed automatically by conda, but the system-specific NVIDIA GPU driver must be installed first.

Conda is a cross-platform open-source suite as well as a setting supervision structure. The original dataset is to have been trained on Google Colab, and the system requirements are Ultralytics YOLOv8.0.87, Python-3.9.16 torch-2.0.0+cu118 CUDA:0 (Tesla T4, 15102MiB).

The augmented data is trained using the local system specification Ultralyt-ics YOLOv8.0.86 Python-3.9.16 torch-2.0.0 CUDA:0 (16 CPUs, 27.9 GB RAM, and 129.4/293.2 GB disc) with NVIDIA GeForce RTX 3060 Laptop GPU.

3.2. Methodology

This study suggests two datasets with labels that have been expanded from 6 to 8 together with the cutting-edge YOLOv8 deep learning method. Before performing image rescaling, the transformation was first used for image augmentation.

In order to distinguish whether workers were wearing helmets or not, a computer vision system was created utilizing the YOLOv8 object detection method. The methodological stages involved in safety helmet identification in this research are shown in Figure 2.

3.3. Data Collection

In the proposed work, the videos and images are collected from both real-life and available open-source sites. Since our focus is to build a multipurpose helmet detection dataset, the data were collected from traffic areas, construction sites, and industrial factories.

The video data is extracted frame by frame at certain intervals which led to lots of extraction of similar types of frames that were deleted manually to select only the best quality images. To reduce the memory requirement the data were resized at 640*640 dimension. Finally, a total of 1569 images for annotations were selected.

3.4. Dataset Preparation

The processes in data preparation include data annotation, label balancing, and data augmentation. Eight categories were assigned, including the following: face, mask, vehicle, a person with Helmet, a person without a helmet, Head, as well as Head with Helmet.

The annotated file has been chosen over other data annotation formats for this study due to its simplicity (one-line normalization per label), small storage need (text format), and rapid calculation for the Yolo model, as the name implies itself. In the proposed work, the Label Img tool was used to label the image.

First, we changed the tool's default setting for the number of classes in our dataset to 8, subsequently provided the image directory for opening and labelled storing locations, annotated the objects that belonged to the classes, and then generated a text file. The text file contains information about the image, including its name, coordinates, size, depth, and location.

3.5. Data Augmentation

To build usable Deep Learning algorithms, the validation error must be reduced with the training error. This is quite doable with data augmentation. The enriched data will lessen the gap between the validation and training sets, as well as any

subsequent testing sets, by reflecting a wider spectrum of potential data points.

Table 1 compares the number of classes in the SHEL5K database, the suggested original database, and the extended database for each class. As exhibited in Figures 3(a), (b), (c), (d), (e), and (f) added five image effects to the original dataset-rainy, foggy, snowy, shadow, and sun flare-using the image albuminization python library in order to broaden the applicability of our mechanism. The combined total of labels, including the original labels, is 72683.

3.6. Data Splitting

In Table 2, divide the original database and supplemented dataset at random into training data as well as validation data, with approximately 80% and 20% of the total photos in each set. Our training set for the experiments consists of 9762 labels with 1256 images, while our validation set consists of 2352 labels with 313 images.

The training set for the enhanced dataset has 58130 labels with 7532 images, and the validation set has 14553 labels with 1882 images. These numbers are both much higher than those for the original dataset.

3.7. Model Building

As shown in Figure 4, the YOLOv8 algorithm architecture is made up of a Backbone, Neck, and Head, as well as Loss in the subsections that follow, with great detail about the design concepts for each element of the building.

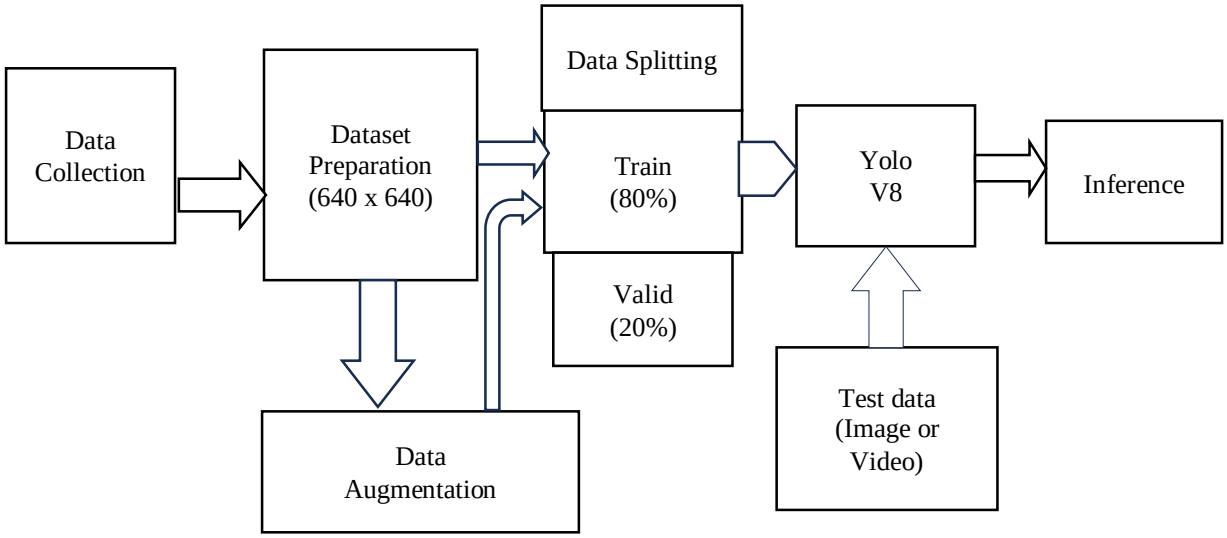


Fig. 2 Schematic workflow of the recommended technique

Table 1. Contrast of the SHEL5K database with the proposed original and augmented dataset

S. No.	Classes	SHEL5K [26]		Original Dataset		Augmented Dataset	
		No. of Labels	Total Images	No. of Labels	Total Images	No. of Labels	Total Images
1	Hat	19252	5000	247	1569	1482	9414
2	Head by hat	16048		1847		11082	
3	Person by hat	14767		1831		10986	
4	Head	6120		2125		12749	
5	Person without hat	5248		2079		12474	
6	Face	14135		1547		9282	
7	Mask	-		477		2862	
8	Vehicle	-		1961		11766	
	Total	75570		12114		72683	

Table 2. Data splitting with respect to original and augmented database with class labels

S. No.	Classes	Original Dataset				Augmented Dataset			
		Train (80%)		Validation (20%)		Train (80%)		Validation (20%)	
		Tags	Pictures	Tags	Pictures	Tags	Pictures	Tags	Pictures
1	Hat	192	1256	55	313	1233	7532	249	1882
2	Head by hat	1491		356		8742		2340	
3	Person by hat	1461		370		8674		2312	
4	Head	1703		422		10255		2494	
5	Person without hat	1716		363		10036		2438	
6	Face	1231		316		7377		1905	
7	Mask	391		86		2279		583	
8	Vehicle	1577		384		9534		2232	
	Total	9762		2352		58130		14553	

(a) **(b)**

(c) **(d)**

(e) **(f)**

Fig. 3 (a) Inventive image, (b) Rainy effect, (c) Foggy, (d) Shadow effect, (e) Snow, and (f) Sun flare effect.

Table 3. Compares the many variations of the YOLO with their different parameters

Model	mAP^val 50-95	Size (Piexts)	Speed A100 Tensor RT (ms)	Speed CPU ONNX (ms)	FLOPs (B)	Params (M)
YOLOv8l	52.9	640	2.399	375.2	165.2	43.7
YOLOv8m	50.2	640	1.83	234.7	78.9	25.9
YOLOv8n	37.3	640	0.99	80.4	8.7	3.2
YOLOv8s	44.9	640	1.20	128.4	28.6	11.2
YOLOv8x	53.9	640	3.53	479.1	257.8	68.2

There are five different variants of the YOLOV8: the YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), as well as YOLOv8x (extra big). Activation Function: Sigmoid Weighted Linear Unit (SiLU), commonly known as the switch activation function, was chosen as one of the hyperparameters for training. - Epoch number: 100, Batch size: eight. The other various parameters of the YOLO variants lie in Table 3.

4. Results and Analysis

4.1. Performance Measures

4.1.1. Precision-Recall Curve

In the suggested study, recall (R), precision (P), F1 score, as well as Mean Average Precision (mAP) were utilized as assessment criteria to carry out a fair contrast among the experimental outcomes of the mechanisms. The accuracy, which is described below, is a measure of the likelihood that the projected bounding boxes will match the actual ground truth boxes.

$$P = \frac{TP}{TP+FP} \quad (1)$$

Where True Negative, True Positive, False Negative, as well as False Positive respectively, stand in for TN, TP, FN, and FP. As seen in the demonstration below, the recall shows the likelihood that ground truth objects will be accurately identified.

$$R = \frac{TP}{TP+FN} \quad (2)$$

4.1.2. F1-Score

Additionally, the F1 measure is mathematically represented as the harmonic mean of the technique's accuracy as well as recall in the equation below.

$$F1 = 2 \times \frac{PR}{P+R} \quad (3)$$

4.1.3. Mean Average Precision (mAP)

The score obtained by contrasting the recognized bounding box to the ground truth bounding box is known as the mean Average Precision (mAP). The following calculation for the mAP is provided if the intersection over the blending score of together cases is 50.0.

Fig. 4 A detailed overview of YOLOv8 architecture

$$mAP = \frac{1}{n}\sum_{i=1}^{i=n} AP_i \quad (4)$$

4.1.4. Intersection over Union (IoU)

A traditional statistic for assessing how well an algorithm performs in object recognition tasks is Intersection over Union (IoU). IoU, which denotes the point at which the produced candidate, as well as the ground truth bounding case, intersects, is the proportion of their overlap as well as the union.

$$IoU = \frac{area(C) \cap area(G)}{area(C) \cup area(G)} \quad (5)$$

4.1.5. Loss

The YOLOv8 method contains reversion as well as categorization branches, and the equation for the categorization branch, which pays BCE Cost, is as follows:

$$loss_n = -w[y_n log x_n + (1 - y_n)\log(1 - x_n)] \quad (6)$$

Here, w denotes the weight, yn is the categorized value, and xn denotes the algorithm model's anticipated value. Additionally, the regression branch makes use of CIoU Loss and Distribute Focal Loss (DFL). DFL's equation is as follows, intending to concentrate on increasing the likelihood of the value around the object y:

$$DFL(S_n, S_{n+1}) = -(y_{n+1} - y)\log(S_n + (y - y_n)\, log(S_{n+1})) \quad (7)$$

Where,

$$S_n = \frac{y_{n+1} - y}{y_{n+1} - y_n},\ S_{n+1} = \frac{y - y_n}{y_{n+1} - y_n}$$

By taking into account the aspect ratio of the anticipation as well as the ground truth bounding box, as illustrated below, CIoU Loss adds an influencing element to DIoU Loss:

$$CIoU_{Loss} = L_{GIoU} + \alpha v - \frac{\rho^2}{c^2} \quad (8)$$

Where GIoU stands for the generalized IoU loss, v is the aspect ratio term, ρ is the distance among the centres of the anticipated as well as ground truth bounding boxes, and c is a normalization factor whose value is dependent on the dimensions of the ground truth bounding case.

4.2. Inference Results

The original database was used to train five different YOLOv8 replicas, each of which was further refined, as seen in Tables 4 through 8 below. The YOLOv8l mechanism trained on the initial database attained the highest mAP0.5 of 0.816. For YOLOv8l, the Mask label achieved the best mAP0.5 of 0.921, while the Helmet and Head classes reported the two worst mAP 0.5 values of 0.769 and 0.675. However, the same model achieved higher accuracy when trained on the augmentation dataset, with mAPs of 0.935 for the Helmet and 0.993 for the Head. The Mean Average Precision (mAP) 50 under intersection over union 50 thresholds, precision (Pr), recall (Re), and Mean Average Precision (mAP (50-95)) where iou sweeps at every 5 steps during training and validation like (50 - 55 - 60) until reaches to 95 steps are the results of the validation of the YOLOv8 algorithm. The 8 classes expanded dataset appears to perform well during the validation phase, which increases the precision and recall score and elevates the mAP. The best mAP is given by the YOLOv8l model, which has an original dataset mAP of 0.91 and an expanded dataset mAP of 0.953.

Table 4. YOLOv8n performance on the original database as well as augmented database

Class	Original Dataset				Augmented Dataset			
	YOLOv8n				YOLOv8n			
	Re	mAP 0.5	Pr	mAP (0.50-0.95)	Re	mAP 0.5	Pr	mAP (0.5-0.95)
Hat	0.765	0.837	0.749	0.608	0.753	0.801	0.832	0.654
Head by hat	0.76	0.817	0.815	0.503	0.857	0.904	0.932	0.683
Person by hat	0.84	0.848	0.836	0.568	0.915	0.961	0.925	0.812
Head	0.582	0.618	0.8	0.301	0.636	0.716	0.883	0.404
Person without hat	0.669	0.764	0.82	0.487	0.779	0.879	0.884	0.646
Face	0.68	0.765	0.822	0.393	0.748	0.847	0.901	0.551
Mask	0.727	0.812	0.945	0.511	0.825	0.907	0.95	0.624
Vehicle	0.801	0.857	0.862	0.616	0.827	0.912	0.877	0.69
Average	0.728	0.79	0.831	0.498	0.793	0.866	0.898	0.633

Table 5. YOLOv8 performance on the original database as well as augmented database

Class	Original Dataset				Augmented Dataset			
	YOLOv8s				YOLOv8s			
	Re	mAP 50	Pr	mAP (0.50-0.95)	Re	mAP 0.5	Pr	mAP (0.5-0.95)
Hat	0.714	0.623	0.637	0.485	0.819	0.911	0.912	0.752
Head by hat	0.76	0.815	0.879	0.513	0.888	0.928	0.959	0.768
Person by hat	0.801	0.84	0.804	0.573	0.964	0.984	0.965	0.892
Head	0.656	0.707	0.868	0.352	0.68	0.77	0.91	0.494
Person without hat	0.711	0.805	0.827	0.507	0.864	0.934	0.908	0.746
Face	0.652	0.816	0.908	0.436	0.806	0.889	0.933	0.669
Mask	0.752	0.786	0.861	0.504	0.873	0.938	0.955	0.718
Vehicle	0.81	0.89	0.856	0.641	0.863	0.941	0.908	0.756
Average	0.726	0.785	0.83	0.502	0.844	0.912	0.931	0.724

Table 6. YOLOv8m performance on the original database as well as augmented database

Class	Original Dataset				Augmented Dataset			
	YOLOv8m				YOLOv8m			
	Pr	Re	mAP 50	mAP (0.50-0.95)	Re	mAP 0.5	Pr	mAP (0.5-0.95)
Hat	0.659	0.607	0.604	0.443	0.899	0.848	0.94	0.798
Head by hat	0.854	0.758	0.818	0.527	0.972	0.906	0.941	0.828
Person by hat	0.814	0.802	0.841	0.582	0.986	0.974	0.992	0.932
Head	0.895	0.653	0.742	0.379	0.934	0.723	0.806	0.569
Person without hat	0.8	0.725	0.802	0.516	0.932	0.887	0.955	0.803
Face	0.91	0.688	0.843	0.453	0.937	0.851	0.924	0.745
Mask	0.832	0.762	0.805	0.505	0.98	0.906	0.952	0.776
Vehicle	0.845	0.832	0.884	0.64	0.926	0.88	0.956	0.802
Average	0.826	0.728	0.792	0.506	0.946	0.872	0.933	0.782

Table 7. YOLOv8l performance on the original database, as well as augmented database

Class	Original Dataset				Augmented Dataset			
	YOLOv8l				YOLOv8l			
	Pr	Re	mAP 50	mAP (0.50-0.95)	Re	mAP 0.5	Pr	mAP (0.5-0.95)
Hat	0.71	0.765	0.869	0.57	0.915	0.853	0.955	0.808
Head by hat	0.871	0.822	0.845	0.537	0.974	0.913	0.957	0.851
Person by hat	0.867	0.837	0.888	0.631	0.983	0.98	0.993	0.951
Head	0.805	0.631	0.675	0.329	0.937	0.74	0.825	0.602
Person without hat	0.87	0.713	0.826	0.568	0.941	0.904	0.965	0.831

Face	0.811	0.687	0.773	0.403	0.96	0.875	0.946	0.788
Mask	0.921	0.755	0.862	0.547	0.986	0.917	0.957	0.802
Vehicle	0.861	0.831	0.892	0.661	0.933	0.895	0.963	0.822
Average	**0.841**	**0.755**	**0.916**	**0.531**	**0.954**	**0.885**	**0.953**	**0.807**

Table 8. YOLOv8x performance on the original database as well as augmented database

Class	**Original Dataset**				**Augmented Dataset**			
	YOLOv8x				**YOLOv8x**			
	Re	**mAP 50**	**Pr**	**mAP (0.50-0.95)**	**Re**	**mAP 0.5**	**Pr**	**mAP (0.5-0.95)**
Hat	0.565	0.577	0.676	0.449	0.87	0.962	0.938	0.829
Head by hat	0.849	0.893	0.888	0.581	0.917	0.962	0.979	0.853
Person by hat	0.866	0.885	0.86	0.644	0.98	0.994	0.98	0.951
Head	0.625	0.704	0.862	0.335	0.745	0.827	0.934	0.602
Person without hat	0.725	0.795	0.807	0.526	0.908	0.964	0.935	0.827
Face	0.71	0.808	0.879	0.421	0.882	0.943	0.949	0.787
Mask	0.83	0.893	0.926	0.553	0.917	0.953	0.959	0.805
Vehicle	0.786	0.945	0.83	0.609	0.893	0.963	0.925	0.825
Average	0.744	0.916	0.841	0.515	0.889	0.953	0.95	0.81

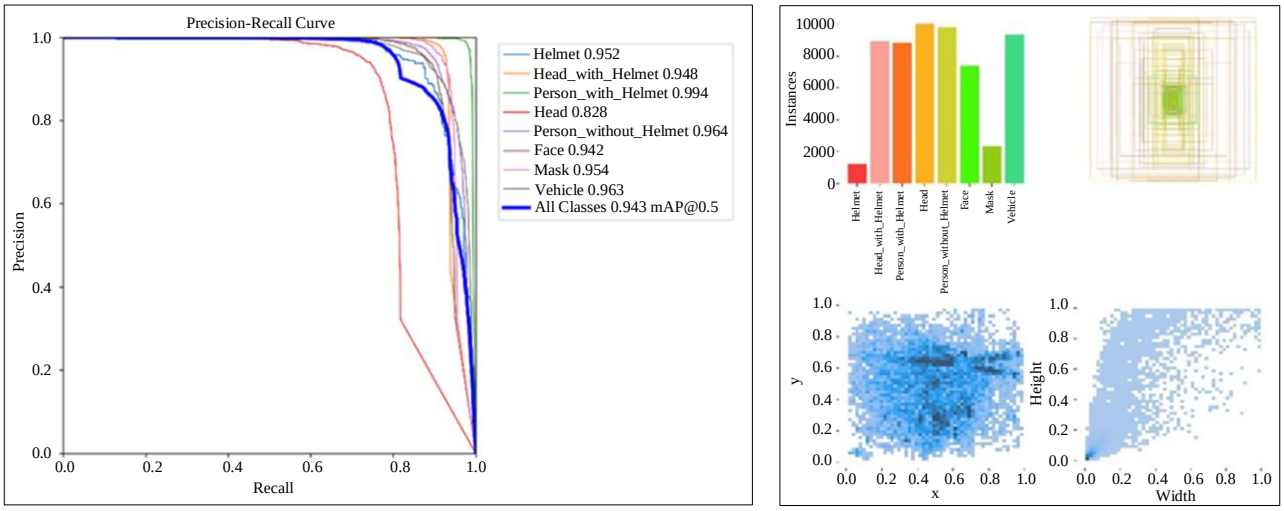


Fig. 5 The YOLOv8x PR curve and label instances on the augmented dataset

F1 scores of each version of the models are illustrated in Figure 6, overall loss, as well as accuracy graphs, are exhibited in Figures 7 (a), and (b) for Yolov8l and the confusion matrix for all the versions is drafted in Figure 8. Table 9 describes the correlation of our suggested dataset to new-fangled models. Table 10 describes the comparison of our suggested dataset to state-of-the-art replicas. The precision, recall, as well as mAP score obtained by the mechanism on the recommended augmented database, were higher than the original database.

The recall, precision, as well as mAP score obtained by the mechanism on the recommended augmented database, were higher than the original database. The recall, precision, as well as mAP of the mechanism on the SHEL5K, were obtained as 0.817, 0.9188, as well as 0.8644, correspondingly.

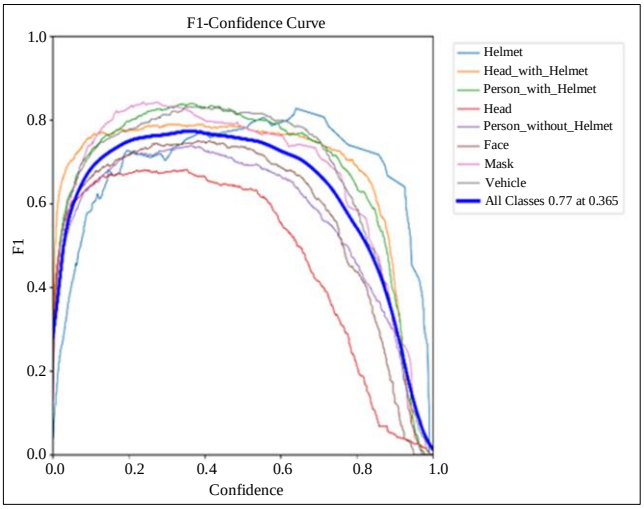
The SHEL5K database has only six labels, while the recommended database has eight labels. Additionally, during the labeling of the recommended database, an image residing

a few portions of the Helmet as well as the face was categorized as the hat or face class, subsequently.

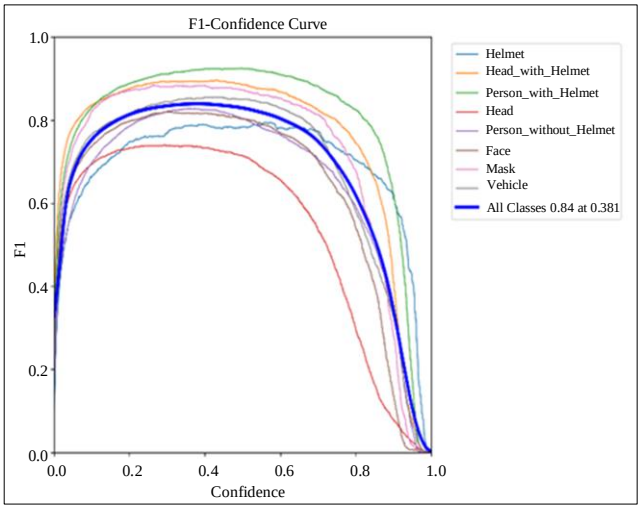
4.3. Prediction on Real-Time Images

The predicted outcomes of the best mechanism YOLOv8l trained on the augmented database as well as the original database on test images were relied on in Figures 9 (a), (b), (c), (d), and (e). The outcome of the technique trained on the original database is illustrated in Figure 9(b), which describes that the mechanism can identify the helmet class precisely, which illustrates that the labelling in the recommended database was executed accurately.

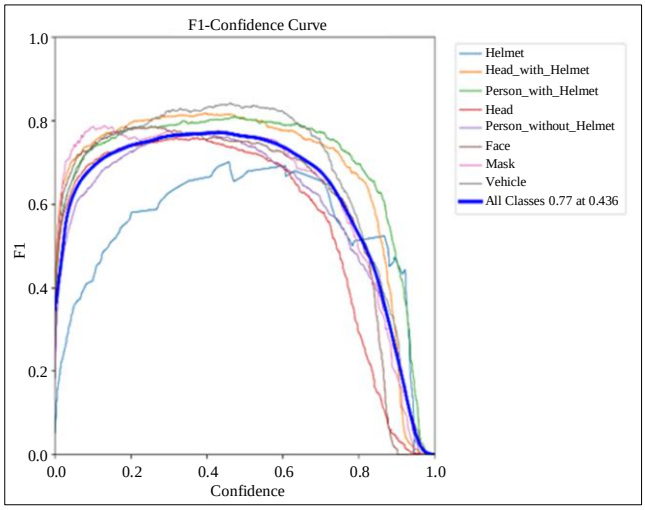
The same helmet database has been employed in this chore to validate the model better. To train helmet detection techniques, the same configuration environment was utilized to train two-level identification technique Faster R-CNN as well as single-stage identification mechanisms FEFDNet, YOLOv7, YOLOv6, YOLOv5s, YOLOv4, YOLOv3, as well as SSD. The contrast between these mechanisms and the YOLOv8l technique relied mostly on AP@50 and mAP@50 as important contrast pointers. Table 10 displays the experiment comparison findings. This suggests that the YOLOv8l algorithm has a greater level of accuracy and performance in place of work safety helmet identification.



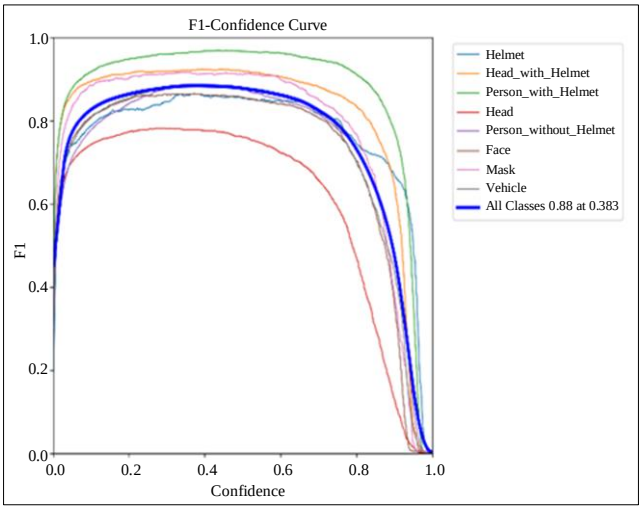
(a) YOLOv8n_original dataset



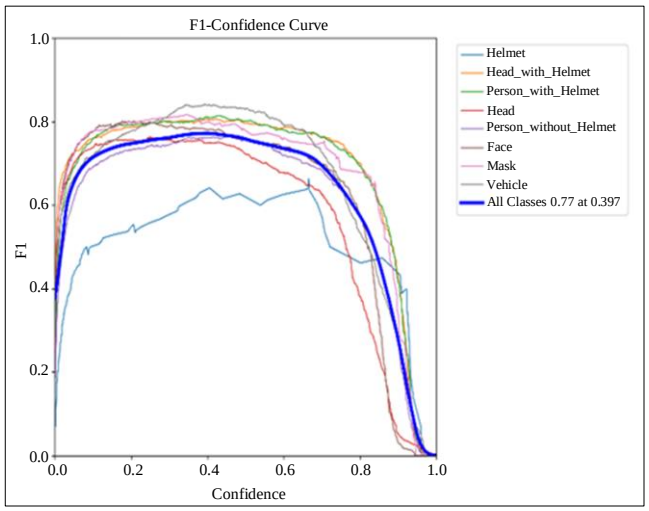
(b) YOLOv8n_augmented dataset



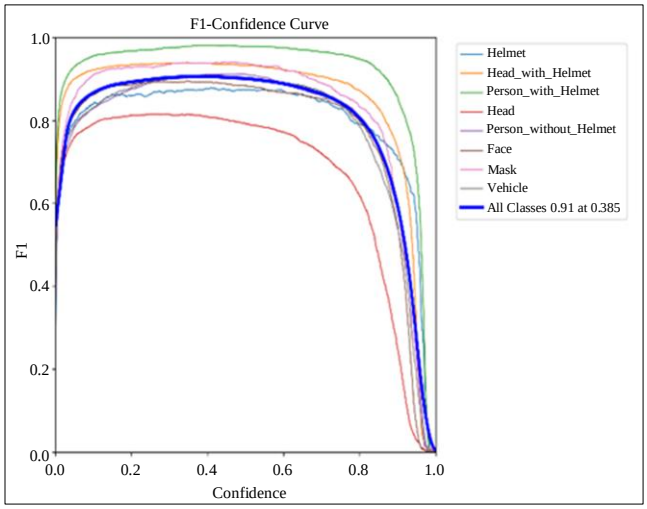
(c) YOLOv8s_original dataset



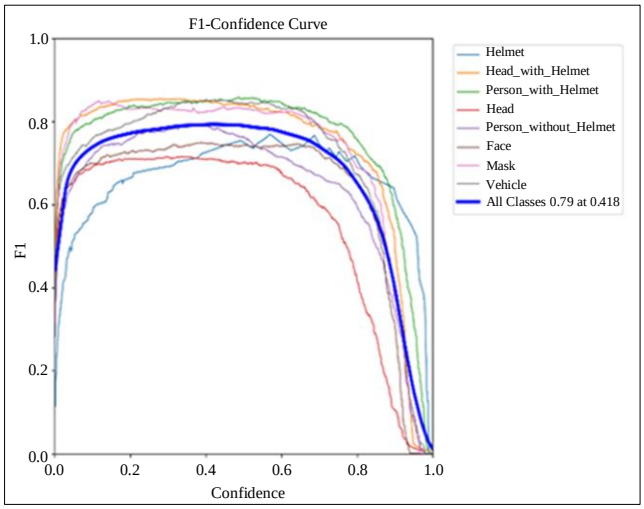
(d) YOLOv8s_augmented dataset



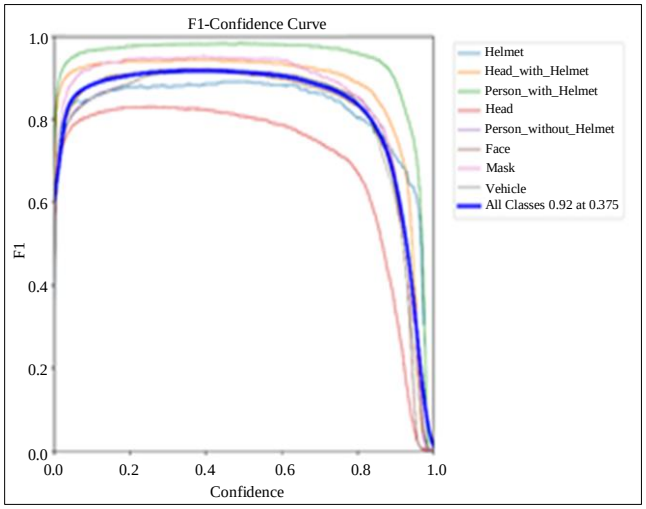
(e) YOLOv8m_original dataset



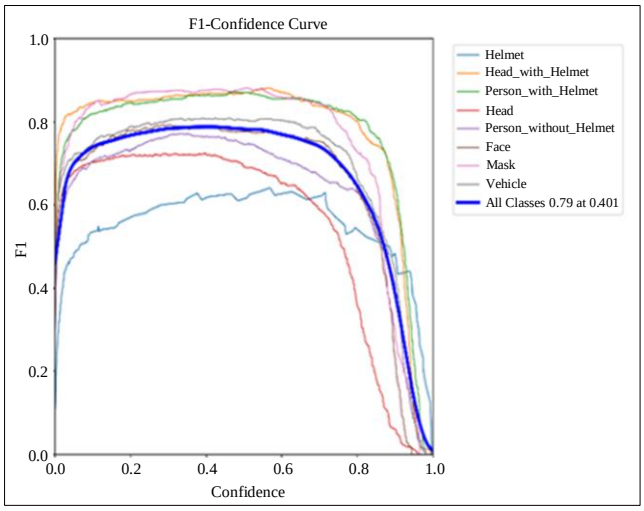
(f) YOLOv8m_augmented dataset



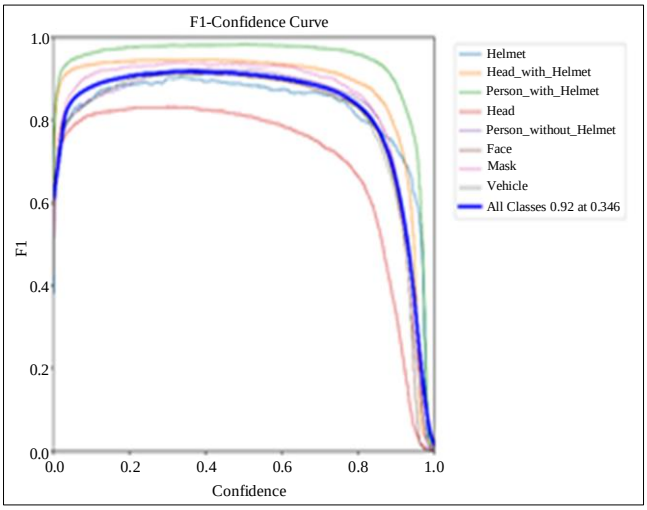
(g) YOLOv8l_original dataset



(h) YOLOv8l_augmented dataset



(i) YOLOv8x_original dataset



(j) YOLOv8x_augmented dataset

Fig. 6 F1 score of YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8m, YOLOv8x respectively (a), (c), (e), (g), (i) on original dataset, and (b), (d), (f), (h), (j) on the augmented dataset.

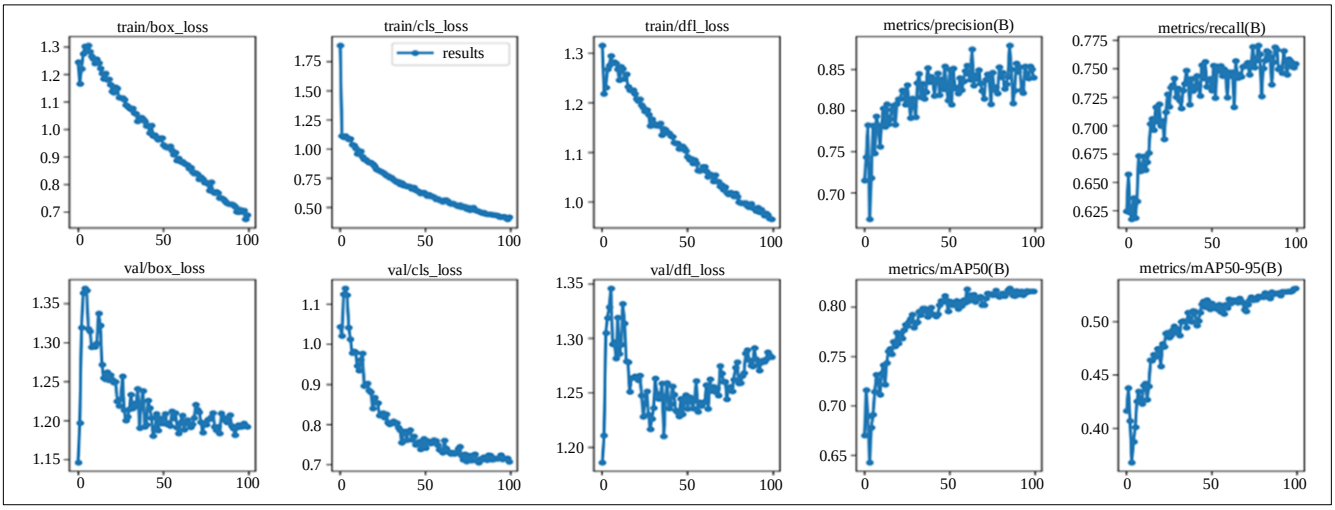


Fig. 7 (a) YOLOv8l_original dataset

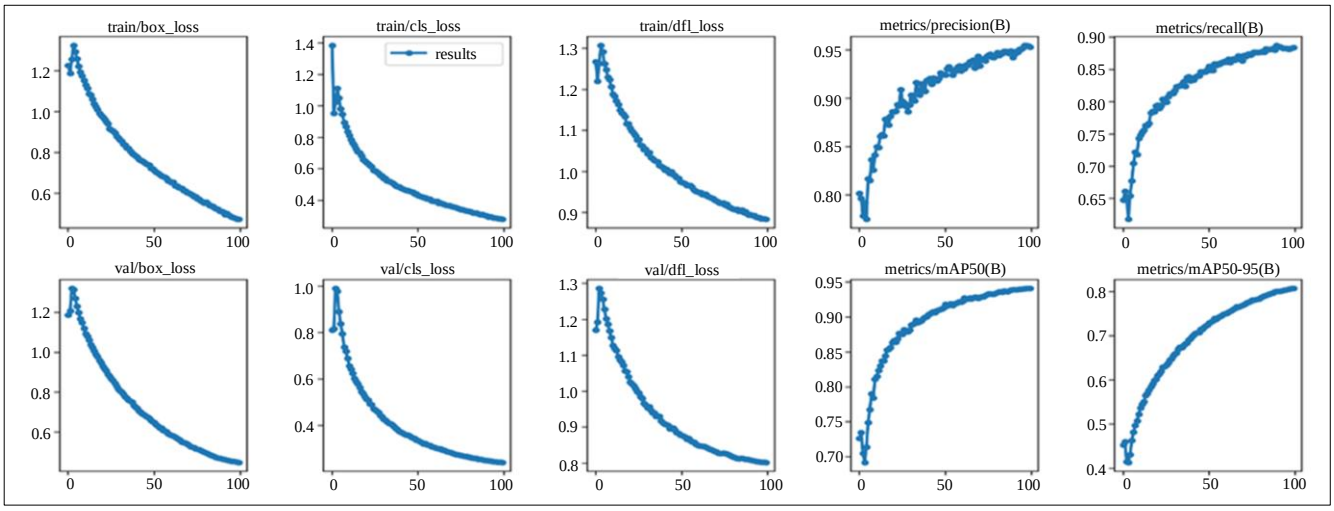
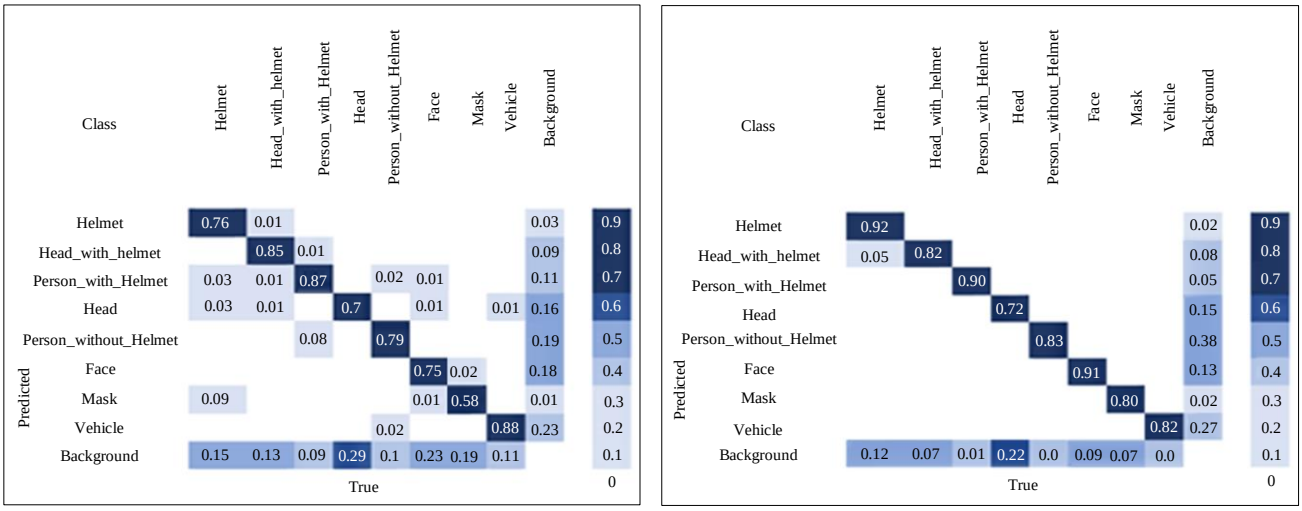


Fig. 7 (b) YOLOv8l_augmented dataset



(a) **(b)**

Fig. 8 Confusion matrix of YOLOv8l (a) Original dataset, and (b) Augmented dataset.

(a) **(b)**

(c) **(d)**

(e)

Fig. 9 Actual prediction with YOLOv8l mechanism by test images taken from the various locations a) Railway station - predicted accuracy with the Helmet is 0.94 mAP, b) Vehicle parking lot - predicted accuracy with the Helmet is 0.95 mAP, c) Road with moving vehicles - predicted accuracy with the Helmet is 0.92 mAP, d) Classroom – predicted accuracy without a helmet is 0.97 mAP, and e) Workplace – predicted accuracy with the Helmet is 0.940 mAP.

Table 9. Comparison results of proposed YoloV8 mechanisms on both the original database as well as the augmented database with different SHEL5K-trained models

Models		Layer	Parameter (million)	Batch Size	Inference Time (ms/img)	Training Time (Hours)	GFL OPs	P	R	mAP 50	F1
Original Dataset	YOLOv8n	168	3.007208	8	5.0	1.734	8.1	0.831	0.728	0.79	0.77
	YOLOv8s	168	11.12868	8	5.3	1.641	28.5	0.83	0.726	0.785	0.77
	YOLOv8m	218	25.844392	8	10.4	1.862	78.7	0.826	0.728	0.792	0.77
	YOLOv8l	**268**	**43.612776**	**8**	**16.8**	**2.22**	**164.8**	**0.841**	**0.755**	**0.916**	**0.79**
	YOLOv8x	268	68.131272	8	25.9	3.316	257.4	0.841	0.744	0.8	0.79
Augmented Dataset	YOLOv8n	168	3.007208	8	5.7	7.0	8.1	0.898	0.796	0.866	0.84
	YOLOv8s	168	11.12868	8	12.9	8.12	28.5	0.931	0.844	0.912	0.88
	YOLOv8m	218	25.844392	8	15.0	10.2	78.7	0.946	0.872	0.933	0.91
	YOLOv8l	**268**	**43.612776**	**8**	**23.8**	**14.0**	**164.8**	**0.954**	**0.885**	**0.951**	**0.92**
	YOLOv8x	268	68.131272	4	95.4	20.45	257.4	0.95	0.889	0.943	0.92
SHEL5K [26]	Fast-R-CNN	48	13.3	16	84.0	55.6	-	0.7808	0.3862	0.3689	0.5167
	YOLOv3-tiny	37	8.7	16	6.0	5.2	-	0.7695	0.4225	0.3779	0.5408
	YOLOv3	222	61.6	16	11.0	24.6	-	0.8509	0.4482	0.417	5848
	YOLOv3-SPP	225	62.6	16	12.0	24.6	-	0.8851	0.5848	0.5572	0.7032
	YOLOv4	488	63.9	16	14.0	11.2	-	0.925	0.7798	0.7693	0.8449
	$YOLOv4_{(pacsp\text{-}x\text{-}mish)}$	488	63.9	16	14.0	14.5	-	0.9195	0.8036	0.7915	0.8567
	YOLOv5s	224	7.1	16	18.0	0.3	-	0.9205	0.774	0.861	0.8397
	YOLOv5m	308	21.1	16	22.0	2.7	-	0.9251	0.7851	0.8687	0.8488
	YOLOv5x	476	87.2	16	32.0	6.3	-	0.9188	0.817	0.8826	0.8644
	YOLOR	665	36.9	16	12.0	9.8	-	0.9322	0.8066	0.8828	0.8637

Table 10. Contrast of outcomes of various state-of-art mechanism

Mechanism	AP-50% (Person)	AP-50% (Helmet)	mAP-50% Average
FEFDNet [24]	94.67%	94.89%	94.78%
YOLOv7 [28]	88.76%	90.67%	89.71%
YOLOv6 [27]	87.51%	90.46%	88.98%
YOLO5s [29]	85.76%	90.26%	88.01%
YOLOv4	85.93%	87.87%	86.90%
YOLOv3	79.68%	81.78%	80.73%
Faster R-CNN	84.55%	82.67%	83.61%
SSD300	74.84%	77.62%	76.23%
YOLOv8l (Proposed)	**97.33%**	**95.50%**	**95.30%**

5. Conclusion

The publicly accessible SHEL5K dataset's six classes were to be increased to eight as part of the proposed study. Several cutting-edge one-stage object identification mechanisms, including YOLOv8l, YOLOv8m, YOLOv8n, YOLOv8s, and YOLOv8x, were benchmarked against the suggested dataset. The experimental findings revealed considerable improvements in the models' mAP0.5l. It is clear from the experimental results of the mechanisms on the suggested database that all of the techniques had good accomplishments in class detection. Additionally, it can be said that the high-dimensional pictures and labelling of the proposed dataset were superior to those of the SHEL5K dataset, as well as state-of-the-art mechanisms.

Additionally, models developed using the suggested dataset may be applied to the real-time identification of safety helmets. As a result, our suggested model presents a new dataset with data augmentation that solely interacts with items of interest to shorten training time and conserve computing resources. A brand-new, cutting-edge Yolov8 DNN model that can precisely recognize the Helmet in any adverse environmental conditions can prevent needless damage and can identify whether the individual is wearing a helmet or not. The proposed dataset may be utilized for additional research studies and can be extended based on the user's requirements. In the forthcoming, we majorly focus on improving the real-time efficacy % of safety hat identification.

Acknowledgment

This work is published with the RDF-Seed fund provided by Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology.

Author Contributions

Dasari Naga Vinod - Formulated, co-written, and co-designed the manuscript.

Partho Adhikari - Performed the simulation work, and co-written the manuscript.

N. Kapileswar, Judy Simon, Phani Kumar Polasi, B. Padmavathi - Formulated and co-written the manuscript.

Data Availability

https://drive.google.com/drive/folders/1nPqtJoLQnvOF-Dept1b_ozsyzMAa9hA0

References

[1] Global Status Report on Road Safety 2018, World Health Organization, 2018. [Google Scholar] [Publisher Link]

[2] WHO Sri Lanka, World Health Organization. [Online]. Available: https://www.who.int/srilanka

[3] Qi Wang, Junyu Gao, and Xuelong Li, "Weakly Supervised Adversarial Domain Adaptation for Semantic Segmentation in Urban Scenes," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4376-4386, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[4] Shaoqing Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[5] Kaiming He et al., "Mask R-CNN," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2961-2969, 2017. [Google Scholar] [Publisher Link]

[6] Joseph Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, 2016. [Google Scholar] [Publisher Link]

[7] Wei Liu et al., "SSD: Single Shot Multibox Detector," *14th European Conference*, The Netherlands, pp. 21-37, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[8] Tsung-Yi Lin et al., "Focal Loss for Dense Object Detection," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2980-2988, 2017. [Google Scholar] [Publisher Link]

[9] Shifeng Zhang et al., "Bridging the Gap between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9759-976, 2020. [Google Scholar] [Publisher Link]

[10] Zhi Tian et al., "FCOS: Fully Convolutional One-Stage Object Detection," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Korea (South), pp. 9626-9635, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ze Yang et al., "Reppoints: Point Set Representation for Object Detection," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9657-9666, 2019. [Google Scholar] [Publisher Link]

[12] ultralytics/yolov5, Github. [Online]. Available: https://github.com/ultralytics/yolov5

[13] Linu Shine, and C.V. Jiji, "Automated Detection of Helmet on Motorcyclists from Traffic Surveillance Videos: A Comparative Analysis Using Hand-Crafted Features and CNN," *Multimedia Tools and Applications*, vol. 79, pp. 14179-14199, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] C. Vishnu et al., "Detection of Motorcyclists without Helmet in Videos Using Convolutional Neural Network," *2017 International Joint Conference on Neural Networks (IJCNN)*, USA, pp. 3036-3041, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[15] B. Yogameena, K. Menaka, and S. Saravana Perumaal, "Deep Learning-Based Helmet Wear Analysis of a Motorcycle Rider for an Intelligent Surveillance System," *IET Intelligent Transport Systems*, vol. 13, no. 7, pp. 1190-1198, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Lokesh Allamki et al., "Helmet Detection Using Machine Learning and Automatic License Plate Recognition," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 12, pp. 80-84, 2019. [Google Scholar] [Publisher Link]

[17] Fahad A. Khan, Nitin Nagori, and Ameya Naik, "Helmet and Number Plate detection of Motorcyclists Using Deep Learning and Advanced Machine Vision Techniques," *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, India, pp. 714-717, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] A. Farhadi, and J. Redmon, "Yolov3: An Incremental Improvement," *Arxiv*, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[19] Apoorva Saumya et al., "Machine Learning Based Surveillance System for Detection of Bike Riders without Helmet and Triple Rides," *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, India, pp. 347-352, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] C.A. Rohith et al., "An Efficient Helmet Detection for MVD Using Deep Learning," *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, India, pp. 282-286, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[21] Aphinya Chairat et al., "Low Cost, High Performance Automatic Motorcycle Helmet Violation Detection," *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 3560-3568, 2020. [Google Scholar] [Publisher Link]

[22] Madhuchhanda Dasgupta, Oishila Bandyopadhyay, and Sanjay Chatterji, "Automated Helmet Detection for Multiple Motorcycle Riders Using CNN," *2019 IEEE Conference on Information and Communication Technology*, India, pp. 1-4, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[23] Hanhe Lin et al., "Helmet Use Detection of Tracked Motorcycles Using CNN-Based Multi-Task Learning," *IEEE Access*, vol. 8, pp. 162073-162084, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[24] Felix Wilhelm Siebert, and Hanhe Lin, "Detecting Motorcycle Helmet Use with Deep Learning," *Accident Analysis and Prevention*, vol. 134, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[25] Wei Jia et al., "Real-Time Automatic Helmet Detection of Motorcyclists in Urban Traffic Using Improved YOLOv5 Detector," *IET Image Processing*, vol. 15, no. 14, pp. 3623-3637, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[26] Munkh-Erdene Otgonbold et al., "SHEL5K: An Extended Dataset and Benchmarking for Safety Helmet Detection," *Sensors*, vol. 22, no. 6, pp. 1-23, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[27] Chuyi Li et al., "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications," *ArXiv*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7464-7475, 2023. [Google Scholar] [Publisher Link]

[29] Yiduo Zhang, Yi Qiu, and Huihui Bai, "FEFD-YOLOV5: A Helmet Detection Algorithm Combined with Feature Enhancement and Feature Denoising," *Electronics*, vol. 12, no. 13, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]