Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

# Real-Time Safety Helmet Detection and Restricted Zone Alert using Enhanced YOLOv8 with Sparse Spatial Attention

**Akash K.**[1]**, Santhi V.**[2]

Computer science and Engineering, PSG College of Technology, Anna University, Coimbatore, India.

**E-mail:** [1]24mz32@psgtech.ac.in, [2]vsr.cse@psgtech.ac.in

## Abstract

Wearing safety helmets is a crucial preventive measure to minimize head injuries in industrial and construction environments preventing unauthorized entry into restricted zones and implementing regulations in helmet wearing are still remain significant challenges in workplace safety management. This work proposes a real-time helmet detection and alert system for the restricted zone based on advanced YOLOv8 developed to improve the accuracy detection. In this model, a Sparse Spatial Attention Mechanism (SSAM) is integrated into YOLOv8 to improve the network focus on required spatial regions by reducing the irrelevant additional features and utilize the modified slim-neck structure. It will be integrated with V0V-GSCSP and GSConv enables efficient multi-scale feature fusion with reduced complexity without the lack of accuracy. This advanced architectural model allows increased precision detection and make it suitable for real-time industrial applications. Additionally, this model improves the real-time alert system developed using python, FastAPI at the backend and Angular at frontend. The system provides the facility to connect an IP camera or web camera as an input source, allows the user to mark restricted zones visually and monitor detections in real time. When any person without a helmet enters the restricted area, an automatic alert message is generated using the Twilio API which instantly send alert message to registered phone number. The enhanced YOLOv8 model achieved better performance with precision of 93.49%, recall of 91.37% and F1-score of 92.40% for the helmet-on class and precision of 92.68%, recall of 91.89% and F1-score of 92.35% for helmet-no class. The overall performances were precision of 93.09%, recall of 91.63% and F1-score of 92.35%. The solution


DOI: https://doi.org/10.36548/jscp.2026.1.004
Received: 14.01.2026, received in revised form: 18.02.2026, accepted: 06.03.2026, published: 19.03.2026

for automated industrial and construction site safety monitoring integrated with slim-neck, SSAM and optimization technique to increase the accuracy detection and reliability.

**Keywords:** Safety Helmet Detection, Restricted Zone Alert, Sparse Spatial Attention Mechanism (SSAM), Slim-neck, GSConv, VoV-GSCSP.

## 1. Introduction

Safety is one of the most critical concerns in industrial and construction environments where workers are frequently exposed to hazardous conditions. Among all safety measures, wearing helmets is one of the most effective ways to protect workers from serious head injuries. However, even with strict rules for safety, ensuring helmet regulations in large and busy worksites remains a problem. Manual supervision is not only labor-intensive but also capable of errors caused by human factors such as fatigue and late response. This work focused on developing a real time monitoring system that can automatically detect whether individuals are wearing helmets and identify unauthorized entries into restricted zones. It has been designed to improve workplace safety through instant alerts in the event of a safety violation. Selecting an appropriate technique for real-time industrial safety monitoring requires consideration of several critical factors:(i) real-time inference speed for continuous video surveillance (ii) robustness in detecting small objects such as helmets and (iii) scalability for integration with alerting and monitoring systems. The enhanced YOLOv8 architecture was selected because it balances detection accuracy, computational efficiency and real-time performance required for industrial safety applications.

### 1.1 Problem Statement

Industrial and construction sites require continuous monitoring to ensure worker safety and prevent unauthorized entries to hazardous or restricted zones. Existing safety supervision approaches mostly depends on manual monitoring and conventional surveillance systems, which are delayed or slow response and human error. Current computer vision solutions lack real-time efficiency, fail to integrate restricted-zone monitoring or not provide immediate alerts when violations occur. As a result, safety violations may unnoticed which increases the risk of workplace accidents. Therefore, a real-time restricted zone monitoring system is required to accurately detect helmet compliances, identify unauthorized entry into restricted areas and generate immediate alert message to safety supervisor

## 1.2 Reason for Choosing the YOLOv8 Model

The enhanced YOLOv8 is chosen for its better performance, overcoming the challenges are observed in real time helmet detection and restricted zone monitoring. Some other models like YOLOv3 and YOLOv5 provide better results, but their accuracy and computation speed was not efficient for real-time situations. YOLOv11 obtained the better results, but it's computationally expensive and it does not suitable for devices with mid specifications systems which is used in the industry. So, YOLOv8 reduced between high accuracy, computation less expensive when compared to YOLOv11 and it is suitable for industrial applications. The YOLOv8 model is enhanced with Sparse Spatial Attention Mechanism (SSAM) to the backbone and slim-neck with GSConv and VoV-GSCSP modules to focus on the important spatial regions, and reduce irrelevant computations during multiple epochs. The AdamW optimizer is used for weight decay results in increased generalization and stabilize the model training and data augmentation for enhanced varied training data by rotating, flipping of images. The enhanced YOLOv8 model is obtained with better results and it will be discussed in following chapters.

## 1.3 Research Contribution

The main contributions of this work are:

- Integration of a Sparse Spatial Attention Mechanism (SSAM) into YOLOv8 to enhance spatial feature or global feature discrimination for helmet detection.
- Design of a Slim-Neck architecture using GSConv and VoV-GSCSP to improve multi-scale feature fusion which reduces computational overhead.
- Development of a restricted-zone intrusion monitoring mechanism for safety enforcement.
- Implementation of a real-time alert system that notifiy safety supervisors immediately via message.
- Demonstrating improved detection accuracy and real-time feasibility for industrial safety monitoring.

## 2. Related Work

The proposed work an enhanced YOLOv8 based helmet detection model called YOLOv8 SLIM-CA [1], this model designed to overcome the limitation of small object detection in complex construction environments. The model use Mosaic data augmentation, Coordinate Attention (CA) and a Slim-Neck structure is used instead of normal neck. For the SHWD dataset, it enhanced mAP by 3.549%, decrease parameters by 6.98%, and reduce computation by 9.76% when compared to YOLOv8. This work developed a lightweight YOLOv5n based model for detecting helmet-wearing behavior in complex conditions [2]. It has Efficient IoU loss, Soft NMS, CAM attention. The proposed work is an enhanced YOLOv5 based algorithm for helmet detection [3]. This is mainly developed to overcome limitation like less performance under low lighting and occlusion, Modification of model included a FasterNet backbone and CAM attention mechanism. The work introduced a framework which is based on YOLOv3 for risk detection and trajectory tracking at construction sites [4]. YOLOv3 is used for helmet and safety vest detection, the system used Kalman filtering and the Hungarian algorithm for efficient tracking. The proposed multi-scale YOLO-based helmet detection for construction safety [5].

This model integrated multi-scale feature learning, the approach enhanced detection robustness across different helmet sizes and construction worker positions and activities. The system uses YOLOv5 for safety helmet detection using the custom SHWD dataset [6]. This model shows better detection accuracy and inference speed. But this model struggles in dense occlusion scenarios, construction workers where overlapping reduces of recognition performance. This method [7] developed YOLOv3 algorithm for detecting multiple safety gears like helmets, gloves and boots in power substation operations and industrial sites. The model integrates gamma correction, K-means box optimization and transfer learning. This make model highly accurate and faster detection when compared to standard YOLOv3. The proposed work of optimized YOLOv5-based approach for safety helmet and flame detection [8], aiming to improve detection efficiency and enable real-time monitoring in complex environments. This approach also introduced optimizations for lightweight deployment, improve detection efficiency while enabling real time monitoring of construction site workers. This work investigated YOLOv5 CBAM-DCN for helmet violation identification model [9]. This model used CBAM attention Mechanism, deformable convolution network (DCN) and Distance IoU loss, which effectively improve feature extraction and detection accuracy. This model is tested on both SHWD and a custom violation dataset, it obtained 91.6% accuracy, a 2.3% improvement over standard YOLOv5. The proposed FEFD-YOLOv5 combining feature enhancement and denoising modules with YOLOv5 to improve robustness under noisy conditions [10].

Using the BJTU Helmet Dataset, the model achieved 94.89% accuracy in noiseless and 91.55% in noisy conditions, outperforming baseline YOLOv5. The initial studies of safety helmet detection were aimed at enhancing safety in construction sites by using automated visual inspection. One of the early hardhat detection systems proposed by [11] aimed at strengthening compliance of workers in construction sites by improving their safety. Their method proved that computer vision could be used to monitor safety, but based on traditional detection methods that were not efficient in hard lighting and conclusiveness conditions. The study [12] established an automated hardhat detection system of construction safety applications based on the traditional image processing and machine learning algorithms. Even though their system enhanced automation of detection, it could not detect dynamically and could not scale in real time. The development of deep learning led to a high detection rate of objects. R-CNN [13].

The research work based on region-based convolutional network and applied rich hierarchy of features to detect and segment objects. This work formed the basis of the present day detection structure but it was costly in terms of computation. This research suggested Fast R-CNN [14] that maximized the sharing of features and training efficiency. It was improved by Faster R CNN [15] which made use of Region Proposal Networks (RPN) to execute in near real time. Although this is highly accurate, these models are still computationally expensive to run in real-time in an industrial setup. In addition to the helmet detection, [16] took the concept of feature fusion of infrared head pose estimation and proved that a multi-scale spatial feature is significant in ensuring a robust human-related detection problem in extreme environmental issues. Recent studies have been concentrating on lightweight deep learning based safety compliance detectors in real time. The research [17] has developed a model of the improved version of the YOLO-based helmet detector which enhanced the accuracy and strength in the areas with the complex conditions. In the same way, this method [18] described a YOLOv5-based framework that integrates hierarchy of positive sample selection and box density filter to enhance efficiency of detection and the choice of false positives. In search of a more efficient deployment, the work [19] introduced a simplistic helmet detecting model that uses an improved architecture of YOLOv4, which can be inferred at a lower rate using less computation effort. Further, the method [20] suggested a multi-scale safety helmet detecting method based on the use of SAS-YOLOv3-tiny, which enhanced the speed of detection and the representation of multi-scales, which is networkable to embedded and edge devices. Table 1 represents the comparative review of existing work.

**Table 1.** Comparative Review of Existing Works

| Approach Category | Representative Methods | Core Characteristics | Strengths | Research Gaps | Ref. |
|---|---|---|---|---|---|
| YOLOv8-based Optimization | YOLOv8 SLIM-CA | Slim-Neck, Coordinate Attention | Improved small object detection, reduced parameters | Limited performance in dense occlusion | [1] |
| Lightweight YOLO Models | YOLOv5n | Efficient IoU, Soft-NMS, CAM | Lightweight and real-time capable | Reduced accuracy in complex scenes | [2] |
| Backbone Enhancement | YOLOv5, FasterNet | Improved feature extraction backbone | Better performance in low-light, occlusion | Increased computational complexity | [3] |
| YOLOv3-based Detection | YOLOv3, tracking | Detection, tracking (Kalman filter) | Real-time tracking capability | Lower accuracy vs modern YOLO | [4] |
| Multi-scale Detection | Multi-scale YOLO | Multi-scale feature extraction | Better detection of varying helmet sizes | Higher computation cost | [5] |
| YOLOv5 Baseline | Standard YOLOv5 | Balanced speed, accuracy | Good real-time performance | Weak under occlusion | [6] |
| Feature Optimization | YOLOv3, preprocessing | Gamma correction, K-means anchors | Improved detection accuracy | Limited adaptability | [7] |
| Multi-task Detection | YOLOv5 (Helmet, Flame) | Multi-object detection system | Real-time monitoring | Reduced task-specific accuracy | [8] |
| Attention-based Models | YOLOv5 CBAM-DCN | Attention, deformable convolution | Better feature extraction | High computation overhead | [9] |
| Noise Robust Models | FEFD-YOLOv5 | Feature enhancement, denoising | Robust in noisy environments | Slight performance drop in extreme noise | [10] |
| Improved YOLO Frameworks | YOLO-based improved model | Enhanced robustness in complex scenes | Better accuracy | Not optimized for lightweight deployment | [17] |

| | | | | | |
|---|---|---|---|---|---|
| False Positive Reduction | YOLOv5, sample selection | Improved positive sample selection | Reduced false detections | Increased training complexity | [18] |
| Lightweight YOLOv4 | Improved YOLOv4 | Reduced computation | Faster inference | Lower accuracy than YOLOv5, YOLOv8 | [19] |
| Edge-Optimized Models | SAS-YOLOv3n | Lightweight for embedded systems | High speed, edge deployment | Accuracy trade-off | [20] |

## 3. Proposed Work

### 3.1 Work Flow of Proposed System

The workflow of the proposed system, as shown in the Figure 1 represents the identification of safety compliance through video analytics. The system integrates computer vision, deep learning and real-time alerting mechanisms to enable safety monitoring at industries or construction sites. It focuses on two major aspects: helmet detection and restricted zone monitoring using a YOLOv8-based customized detection model combined with backend automation and live video streaming. Video Input Acquisition is the system begins with the acquisition of live video feed serves as the input source for the detection pipeline. It can be sourced from a webcam or an IP camera URL depending on the environment of deployment. The collection and processing of these video frames in real time are done at the backend, develop using FastAPI with the OpenCV library. Every frame is effectively sent to the frontend using WebSocket communication for real time visualization. This system allows operators to monitor live activity to detect area with minimal latency. Before detection select the restricted zone, the supervisor or safety officer first defines a restricted zone precisely termed as 'danger area', directly through the front-end user interface. This is implemented by creating a rectangular region on the live video stream to identify the area that requires restricted access. The coordinates are captured and securely transmitted to the backend. Now, the backend continuously compares the location of the detected person with the defined area to find out if the person is within the restricted zone or not. This feature ensures dynamic adaptability of different zones can be defined for various operational sites.

Object Detection using enhanced YOLOv8 is developed in each frame with the help of a custom YOLOv8 model when the frames are ready. Added to this model, a custom-designed SSAM increases the network to pay more attention to the main regions that the head and upper body of the worker. Slim Neck architecture has also been enhanced to achieve lesser computational complexity, this makes the network faster without lack in detection performance. Helmet and zone compliance check, if a person detected with a safety helmet outside the restricted zone, then it is a safe condition and it continue monitoring without any alarms. In any case, upon detection of person inside the restricted zone without wearing safety helmet, it will be treated as violation. This auto-classification will help in appropriate detection of unsafe entry into the specified danger zone. In the case where a worker enters a restricted zone without wearing a helmet, the system flags the violation as more serious, the backend would initiate an alert system and sent immediate notifications to the registered mobile using Twilio API.

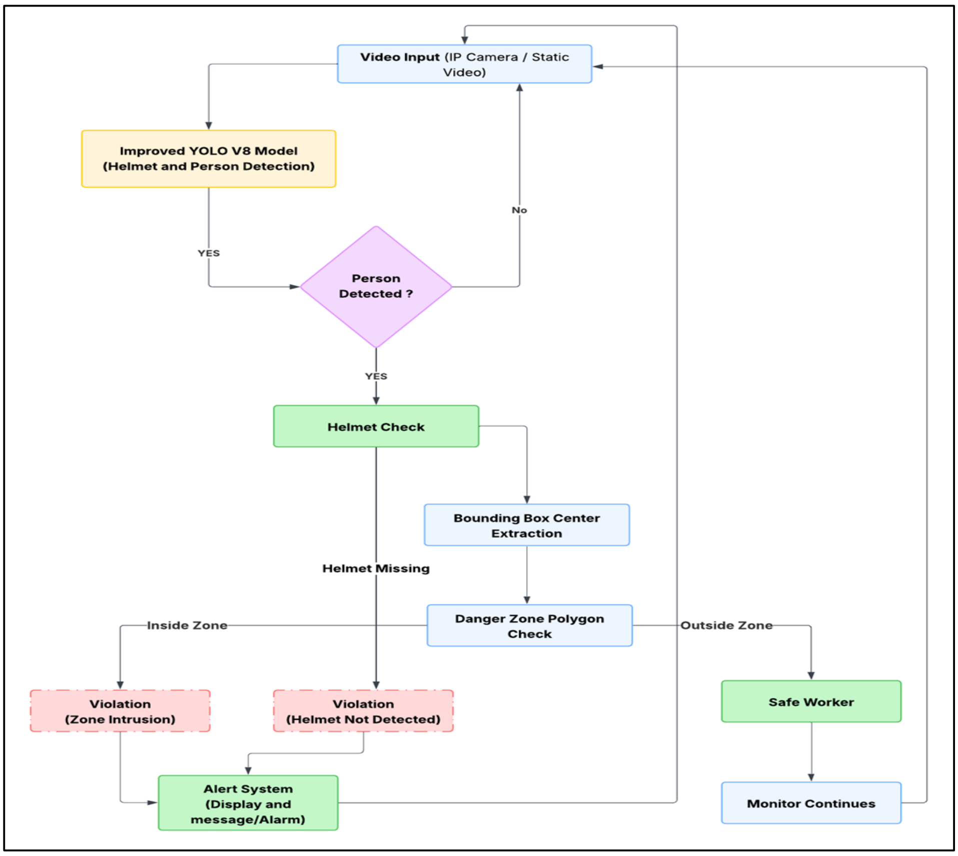


**Figure 1.** Overall Functional Elements of the Proposed System

### 3.2 Overview of Enhanced YOLOv8 Model

The proposed system developed on the basis of the YOLOv8 object detecting architecture that is widely known for its low weight and high real-time performance. The default YOLOv8 model can be constrained in complicated industrial scenes where safety helmets are small values partially occluded, or under overloaded backgrounds. To overcome these limitations, the improved YOLOv8 architecture is introduced using a Slim-Neck design based on GSConv and VoV-GSCSP modules and a Sparse Spatial Attention Mechanism (SSAM) architecture used in the backbone. To improve the representation of spatial features as it concentrates on discriminative features of the workers like the head or helmet part and rejects unnecessary background data. Moreover, the Slim-Neck architecture enhances multi-scale features fusion and minimizes the computational complexity makes it effective in real-time detection. These architectural changes improve the capacity to extract features and high levels of computational efficiency without necessitating many new model parameters. The enhanced YOLOv8 was proposed and trained on the safety helmet dataset and measured for conventional object detection indicators. The results of the experiment prove that the superior architecture can achieve a precision of 0.9309, a recall of 0.9163 and F1-score of 0.9235. Additionally, the model had a mAP 0.5 of 0.9450, mAP 0.5-0.95 of 0.6789 is better than the baseline YOLOv8 model with a precision of 0.8904, 0.8497 and a mAP 0.5 of 0.9079. These enhancements indicate that the addition of SSAM can considerably increase the spatial attention and detection accuracy, especially of small and partially covered helmet objects. The high-performance in detection with enhanced construction stores real-time inference about 30-45 FPS is appropriate in the use of the enhanced construction in industrial safety monitoring systems.

### 3.3 Dataset Description

This improved model was trained, validated and tested using a custom dataset based on the Safety Helmet Wearing Dataset. About 2500 safety monitoring annotated images are included in the dataset. In this dataset, three classes were taken into account namely, Helmet On, No Helmet and Person. The data was subdivided into training, validation and testing groups in the proportion of 7:2:1 and the result was a ratio of 1750 training images, 500 validation images and 250 testing images**.**

### 3.4 Enhanced YOLOv8 Model Architecture

#### Backbone Architecture

Figure 2 shows enhanced YOLOv8 backbone integrates a sparse-spatial attention mechanism to enhance feature representation by focusing on salient regions while reducing computational redundancy. The advantages include enhanced detection accuracy for complicated areas and better handling of occluded or small objects through prioritized spatial data processing.

**Convolution Layer:** In a convolutional layer (the backbone network's base layer), convolution operations are performed to extract spatial representations of an input image like corners, edges and textures. Each Conv block usually consists of a convolutional layer followed by a batch normalization and an activation function (like SiLU) that stabilize the training process, introduce non-linearity and increase the featured representation.

**Spatial Pyramid Pooling Fast (SSPF) Layer:** The SPPF layer is an enhanced version of the SPP layer used in some of YOLO's earlier iterations. SPPF applies several pooling operations with different kernel sizes to the same feature map to extract multi-scale context. This requires little additional computation compared to customary convolutions saving computational resources.

**Sparse Spatial Attention Mechanism (SSAM) Layer:** The Sparse Spatial Attention Mechanism (SSAM) module is a recently proposed module added to the backbone of the enhanced YOLO-based architecture to allow spatial information about the most relevant regions in the input image. However, in most cases, background regions and redundant textures introduce noise into the input, degrading the model performance. The SSAM reduce this issues by understand the discriminative spatial regions with high attention weights and perform sparse attention to the irrelevant ones. SSAM uses the spatial feature maps produced after passing through convolutional layers. The model computes attention weights across the height and width spatial dimensions, indicating the importance of each spatial position (height, width pair). These weights are learned adaptively during training. Unlike dense attention that computes attention weights for all pixel positions, SSAM exploits the idea of sparse attention on a subset of salient regions in the input, resulting in lower computational costs while maintaining the focus on the most relevant contents of the image. The sparse attention mask is applied element-wise, while increasing the salient image background. In the enhanced backbone, SSAM is inserted after major convolutional blocks, following the Spatial Pyramid Pooling (SPPF) layer. This placement allows the attention module to act on high level, semantically additional features. SSAM helps the network to adaptively rectify these non-ideal features, before passing them onto the detection head. This is important for industrial safety tasks.

Let the input feature map extracted from the backbone network be represented as:

$$F \in \mathbb{R}\hat{}(H\times W\times C) \quad (1)$$

where F denotes the feature map with height H, width W and C channels.

The spatial attention map is computed by applying convolution over the average-pooled and max-pooled feature maps followed by a sigmoid activation function:

$$A_S = \sigma(f_{conv}(AvgPool(F)) + f_{conv}(MaxPool(F))) \quad (2)$$

where σ denotes the sigmoid activation function, fconv represents the convolution operation and as is the resulting spatial attention map

The sparse spatial attention mechanism enhances important spatial regions by applying a sparse mask to the feature map:

$$F' = F \odot M_S \quad (3)$$

where $M_S$ denotes the sparse spatial mask highlighting important regions, $\odot$ represents element-wise multiplication, and $F'$ is the refined feature map.

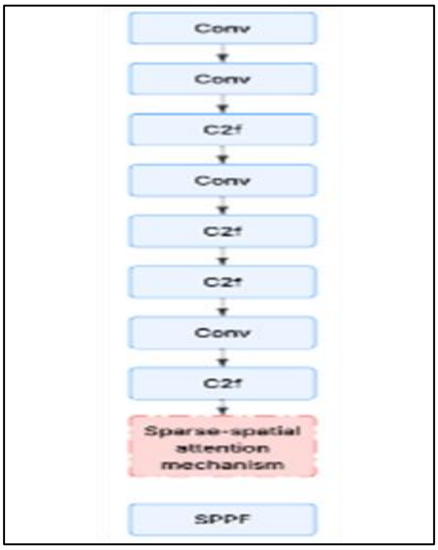


**Figure 2.** Enhanced YOLOv8 Backbone Architecture

**Neck Architecture**

Figure 3 YOLOv8 neck utilizes a VoV-GSCSP and GSConv structure to enhance gradient flow and data fusion between feature maps. Its primary advantage is achieving a superior balance between accuracy and computational efficiency, making it highly suitable for real-time applications on devices with limited resources.

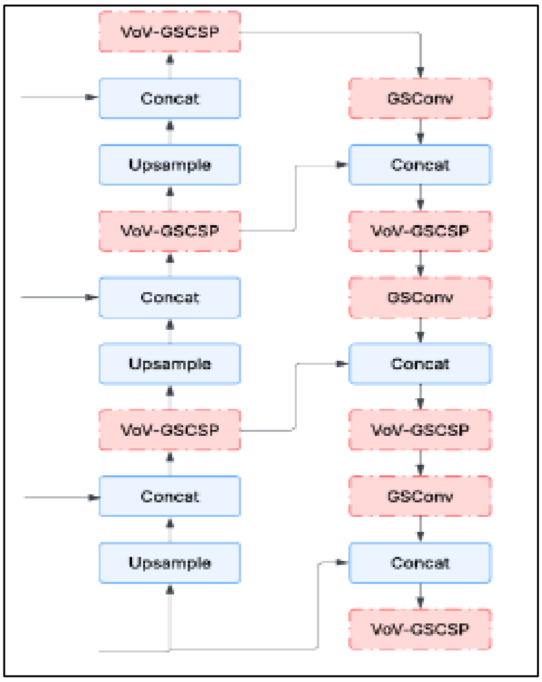


**Figure 3.** Enhanced YOLOv8 Neck Architecture

**VoV-GSCSP Layer:** The VoV-GSCSP (VoVNet Ghost Spatial Cross Stage Partial) block is used for effective efficiency in computation and feature reusability while contributing to gradient flow. It consists of a VoVNet block which can handle multiple receptive fields and a CSP (Cross Stage Partial) block which splits feature maps and merged them to improve gradient flow and reduce parameters. The VoV-GSCSP block is a hybrid module which consists of VoVNet, Ghost Shuffle Convolution (GSConv) and Cross Stage Partial (CSP) module to aggregate. The fundamental building block for the VoV-GSCSP is the OSA (One shot Aggregation) which concatenate the results from several convolutional layers into a single feature map. When compared to standard convolutions, the GSConv provide enhanced feature reuse, special context and reduce computational cost and model parameters. The CSP mechanism splits the input feature map into two main parts: one will process through the VoV-GSConv block and another bypasses processing from the VOV-GSConv block. Eventually, the two results will be merged again to improve gradient flow and reduces unwanted computations. Altogether, the VoV-GSCSP block has better performance between accuracy and efficiency. This will improve overall performance, prevent overfitting and collects multi-scale features effectively.

**GSConv Layer:** GSConv (Ghost Shuffle Convolution) is a lightweight convolutional operation that can effectively lower computational cost providing high representational power.

ISSN: 2582-2640 

GSConv reduces the operation to two parts. The first part generates true features with partial standard convolution and the second, lower cost part generates additional "ghost" features with cheap linear transformations. These operations lead to a substantial decrease in the total number of floating point operations (FLOPs). These characteristics makes GSConv useful for object detection methods in real-time systems and on edge devices.

### 3.5 Performance Comparison of YOLOv8 with Different Attention Mechanism

Table 2. presents the performance comparison of different attention mechanisms integrated with the YOLOv8 architecture. The results indicate that the proposed Sparse Spatial Attention Mechanism (SSAM) achieves the highest detection performance across all evaluation metrics. Specifically, the enhanced YOLOv8 + SSAM model achieves a precision of 0.9309 and recall of 0.9163 outperforming conventional attention modules such as SE, CAM. Furthermore, the proposed model achieves the highest mAP@0.5 of 0.9450 and mAP@0.5–0.95 of 0.6789. This improvement demonstrates that the SSAM module effectively enhances spatial feature representation while suppressing irrelevant background data, leading to more accurate helmet detection in complex industrial environments.

**Table 2.** Performance Comparison of Attention Mechanisms

| Model | Precision | Recall | mAP@0.5 | mAP@0.5–0.95 |
|---|---|---|---|---|
| YOLOv8 | 0.8904 | 0.8497 | 0.9079 | 0.5079 |
| YOLOv8 and SE | 0.9023 | 0.8615 | 0.9187 | 0.5324 |
| YOLOv8 and CAM | 0.9215 | 0.8860 | 0.9314 | 0.5962 |
| YOLOv8 and SSAM | 0.9309 | 0.9163 | 0.9450 | 0.6789 |

Table 3. demonstrates computational efficiency analysis of the proposed enhanced YOLOv8 model achieves superior performance compared to existing models. Although additional modules such as SSAM are integrated into the architecture, the lightweight Slim-Neck structure ensures that the overall computational cost remains low. The proposed model achieves the highest detection accuracy while maintaining real-time processing capability with approximately 45 frames per second, making it suitable for deployment in industrial surveillance systems.

**Table 3.** Computational Efficiency Comparison

| Model | Parameters (M) | FLOPs (G) | FPS |
|---|---|---|---|
| YOLOv3 | 61.5 | 65.9 | 22 |
| YOLOv5 | 7.2 | 16.5 | 35 |
| YOLOv8 | 3.2 | 8.7 | 41 |
| YOLOv8 + CAM | 3.4 | 9.2 | 38 |
| YOLOv8 + SSAM | 3.3 | 8.6 | 44 |

### 3.6 Heatmap Comparison of YOLOv8 with Different Attention Mechanism

Attention heatmaps of the same input image were created with the help of Grad-CAM to determine the effectiveness of specific attention mechanisms. The visualization shown in Fig. 2 highlights the frequency regions to which the model highlights the process of detecting a helmet as shown in Figure. 4. The activation regions in the baseline visualization using the SE attention mechanism (Fig. 4(b)) are relatively spread throughout the image indicating that the model still focuses on the various background areas besides the helmet area. Attention incorporation of CAM (Fig. 4(c)) provides a better concentration on the area of the head to the worker, however, there are still some unwanted activations in the areas. The suggested SSAM-based model (Fig. 4(d)) yields fewer and more localized activation within the area of the helmet in addition to significantly reducing the background response which suggests the Sparse Spatial Attention Mechanism helps to enhance the discriminative spatial features relevant to helmet detection. The comparison of the heatmaps shows that, in comparison to the original design, the YOLOv8 + SSAM architecture provides a better spatial localization of the safety-critical zones which allows the model to focus better on the areas of helmets even in a complex industrial environment. This improved spatial attention is useful in improving the high detection in the proposed model.

Attention heatmaps of the same input image were created with the help of Grad-CAM to determine the effectiveness of specific attention mechanisms. The visualization shown in Fig. 2 highlights the frequency regions to which the model highlights the process of detecting a helmet as shown in Figure. 4. The activation regions in the baseline visualization using the SE attention mechanism (Fig. 4(b)) are relatively spread throughout the image indicating that the model still focuses on the various background areas besides the helmet area.


**Figure 4.** Comparison of Attention Heatmaps Generated by Different Attention Mechanisms. (a) Original Input Image (b) YOLOv8 & SE heatmap (c) YOLOv8 & CAM heatmap (d) YOLOv8 & SSAM heatmap (proposed method)

Attention incorporation of CAM (Fig. 4(c)) provides a better concentration on the area of the head to the worker, however, there are still some unwanted activations in the areas. The suggested SSAM-based model (Fig. 4(d)) yields fewer and more localized activation within the area of the helmet in addition to significantly reducing the background response which suggests the Sparse Spatial Attention Mechanism helps to enhance the discriminative spatial features relevant to helmet detection. The comparison of the heatmaps shows that, in comparison to the original design, the YOLOv8 + SSAM architecture provides a better spatial localization of the safety-critical zones which allows the model to focus better on the areas of helmets even in a complex industrial environment. This improved spatial attention is useful in improving the high detection in the proposed model.

### 3.7 Restricted Zone Selection

The system has an interactive restricted zone selection option which gives the safety supervisor the option to designate a monitoring area directly on the live video feed. User can create a rectangular area as indicated by the green dashed box as illustrated in Figure. 7 around the video frame using the mouse and which is visually indicated by a green dashed box labelled as the Detection Zone. The chosen area is reflective area where monitoring the compliance with safety is conducted on a proactive basis. The frontend interface captures the coordinates of the drawn region and sends the data to the backend server in the form of API communication. After defining the area, the system constantly analyses the objects are observed in this area with the help of the improved YOLOv8 model. In case any individual gets into the restricted area without wearing a safety helmet, the system recognizes the infraction and the alert mechanism is activated. This zone-selection ability is flexible and enables the identification of different monitoring zones on various industrial situations enhancing the flexibility and provision of specific safety surveillance.

**Restricted Zone Representation**

In the proposed system, the restricted zone is defined as a rectangular region selected by the user on the video frame. The region is represented using the coordinates of its top-left and bottom-right corners

$$Z = (x_{min}, y_{min}, x_{max}, y_{max}) \tag{4}$$

where $x_{min}$, $y_{min}$ represent the coordinates of the top-left corner and $x_{max}$, $y_{max}$ represent the coordinates of the bottom-right corner of the restricted zone.

**Centroid Based Calculation of the Detected Person**

For each detected person, the object detection model provides the bounding box coordinates. The centroid of the detected bounding box is calculated to determine the position of the person in the frame.

$$C_x = (x_{min}^b + x_{max}^b) / 2 \tag{5}$$

$$C_y = (y_{min}^b + y_{max}^b) / 2 \tag{6}$$

where $x_{min}^b$, $y_{min}^b$, $x_{max}^b$, $y_{max}^b$ represent the bounding box coordinates of the detected person, and $C_x$, $C_y$ denote the centroid location of the detected object.

**Condition for Person Inside Restricted Zone**

The system determines whether a detected person is inside the restricted zone by checking if the centroid of the bounding box lies within the defined region

$$x_{min} \le C_x \le x_{max} \tag{7}$$

$$y_{min} \le C_y \le y_{max} \tag{8}$$

If both conditions are satisfied, the detected person is considered to be inside the restricted zone.

ISSN: 2582-2640

### Safety Violation Detection

A safety violation is triggered when a person enters the restricted zone without wearing a safety helmet

$$V = (\text{InsideZone} = 1) \wedge (\text{HelmetDetected} = 0) \tag{9}$$

where V represents the violation event, InsideZone indicates whether the person is inside the restricted area, and HelmetDetected indicates whether a safety helmet is detected.

## 3.8 Optimization Technique

AdamW Optimizer is used for the optimization of the enhanced YOLOv8 model parameters. It decouples the weight decay from the gradient-based update provides better generalization and convergence stability. The initial learning rate was 0.001 with a weight decay of 0.0005 to prevent overfitting by penalizing large weights. A cosine learning rate scheduler was adopted, gradually decreasing the learning rate during training, which helps to converge with minimum. In addition, a heat phase of three epochs is applied to stabilize the early training by avoiding sudden large-weight updates. AMP increases computational efficiency and reduces GPU memory utilization for faster training without degradation in accuracy. The early stopping mechanism with a patience of 15 epochs was also applied to prevent over-training once the performance reached a plateau. Several data augmentation strategies were implemented on the training dataset to improve the robustness and generalization capability of the model. Specifically, hue (0.015), saturation (0.7) and value (0.4) in the HSV color space were used to handle illumination changes. The geometric transformations include rotation (±10°), translation (0.1) and scaling (0.5) to further increase the diversity of object positions and sizes in an image. A horizontal flip with a probability of 0.5 was implemented to avoid directional bias in the dataset. A weight decay value of 0.0005 was used to regularize the model by penalizing large weight magnitudes, thereby reducing overfitting and improving generalization. This value falls within the commonly recommended range ($1\times10^{-4}$ to $5\times10^{-4}$) for deep learning models and provide stable convergence and consistent validation performance during training.

## 4. Results and Discussion

Training and validation loss curves in Figure 5 shows a consistent downfall, this indicates an effective model learning and convergence. Both classification and box losses gradually decrease over epochs, results in improved prediction accuracy. The smooth alignment between training and validation losses represents minimal overfitting. Overall, the model performs stable and efficiently optimized throughout the training process.

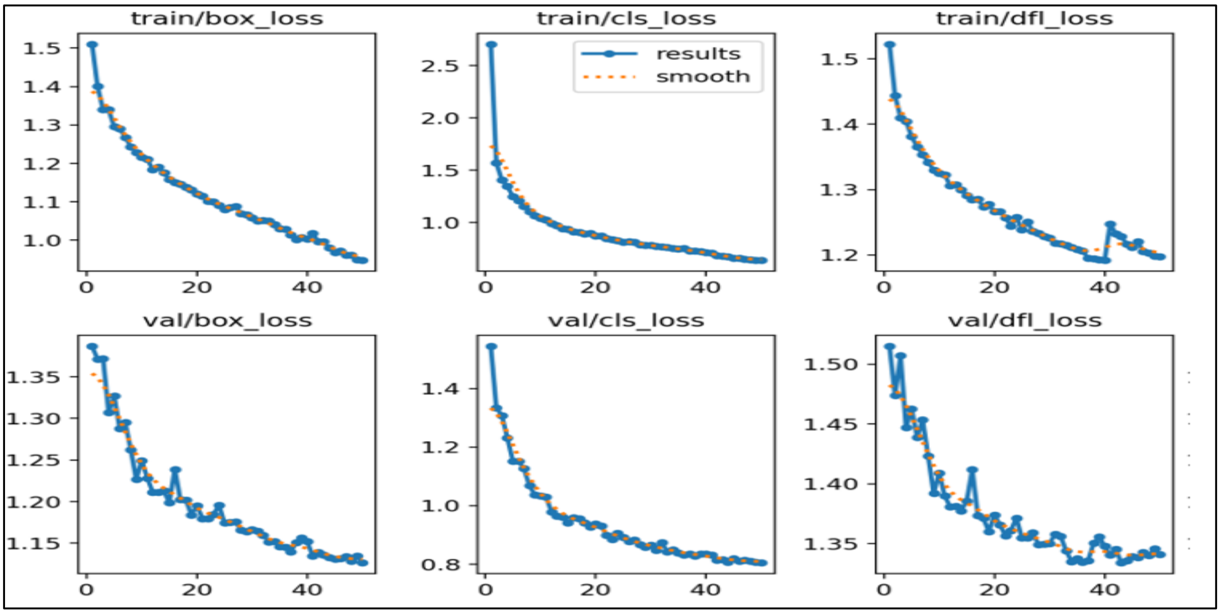


**Figure 5.** Training and Validation Loss Graphs for Enhanced YOLOv8 Model

Figure 6 indicates a confusion matrix of three important classes which included helmet on, no helmet and person that were utilized to safety helmet detection. The matrix has substantial diagonal domination. accuracy of the classifications was high and the discrimination between the classes was better. The classifier was able to recognize and identify 250 instances of Helmet On and this indicates an ability to identify compliant workers. To detect safety-violation, 76 No Helmet cases had been accurately identified, which proves that the model is effective in detecting non-compliant workers in risky situations. Moreover, the system accurately identified 77 Person cases, which demonstrates that the human presence is better before helmet compliance. There were some misclassifications. A small portion of the samples of the Helmet On was estimated as No Helmet (3) or Person (3) and this can be possibly explained by occlusion, motion blur or incorrect positioning of helmets. There is also a level of confusion between No Helmet and Helmet On categories that indicate visual similarity in certain circumstances. The false analyses of Person instances into instances of a helmet are low and could occur when helmets or areas of the head dominate the bounding-box area. To conclude, the confusion matrix shows strong classification results with accuracy of about 92%. The elevated true-positive scores in all three categories confirm the validity of the proposed system of the real-time monitoring of helmet compliance and the implementation of the safety measures of restricted-zones.

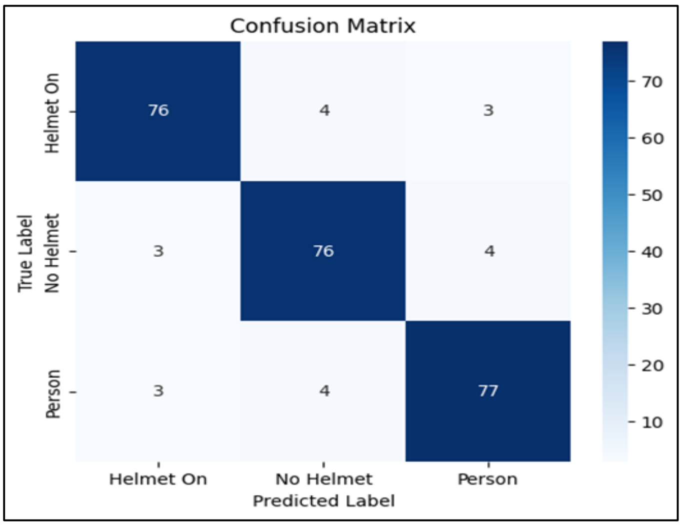


**Figure 6.** Confusion Matrix for Enhanced YOLOv8 Model

Table 4 shows that the enhanced proposed model is very high in accuracy and reliability for Safety helmet detection. The enhanced YOLOv8 model achieved an overall precision of 93.09%, a recall of 91.63% and an F1-score of 92.35% is consistent and stable detection performance. These high precision and recall indicate the effective reduction of false detections in the model while correctly identifying helmet equipped and not helmet-equipped persons.

**Table 4.** Metrics Obtained by Proposed Work

| Metrics | Helmet On | No Helmet | Person | Overall |
|---|---|---|---|---|
| Precision | 93.49% | 92.68% | 93.10% | 93.09% |
| Recall | 91.37% | 91.89% | 91.63% | 91.63% |
| F1-score | 92.40% | 92.31% | 92.34% | 92.35% |

Most importantly, the integration of the SSAM mechanism and AdamW optimizer greatly enhances feature extraction, regularization and significantly improves convergence speed. This model achieved a higher mAP@50 of 94.50% and mAP@50–95 of 67.89%.

Overall, these results show that the proposed model has met the requirements of real-time industrial safety monitoring because of its improvement in accuracy and computational efficiency.

### 4.1 Model Comparison

**Table 4.** Model Comparison Between YOLOv3, YOLOv5, YOLOv8, Enhanced YOLOv8 Over 50 Epochs

| Metrics | YOLOv3 | YOLOv5 | YOLOv8 | YOLOv8 and CAM | YOLOv8 and SSAM | Enhanced YOLOv8 |
|---|---|---|---|---|---|---|
| Precision | 0.8467 | 0.8842 | 0.8904 | 0.9290 | 0.9298 | 0.9309 |
| Recall | 0.7491 | 0.7928 | 0.8497 | 0.8665 | 0.8955 | 0.9163 |
| F1-Score | 0.7949 | 0.8360 | 0.8846 | 0.8966 | 0.9078 | 0.9235 |
| mAP@50 | 0.8232 | 0.8605 | 0.9079 | 0.9314 | 0.9370 | 0.9450 |
| mAP@50-90 | 0.4653 | 0.4983 | 0.5079 | 0.5962 | 0.6115 | 0.6789 |

For performance evaluation, all models were trained using the same dataset, identical training validation splits, equal training epochs and consistent evaluation metrics. This ensures that performance differences increase from architectural improvements rather than training variations. Table 4 presents the performance evaluation results of YOLOv3, YOLOv5, YOLOv8, proposed model (Enhanced YOLOv8). This technique has better performance results when compared to other models, it obtained highest precision at 0.9309, recall at 0.9163 and F1-score at 0.9235 proves its detection precision and stability superiority. The integration of the Sparse Spatial Attention Mechanism (SSAM) into the YOLOv8 backbone combined with the AdamW optimizer produces better mAP@50 results at 0.9450 and mAP@50–95 results at 0.6789 compared to YOLOv3 (0.8232, 0.4653) and YOLOv5 (0.8605, 0.4983). This improvement proves the effectiveness of the proposed approach. The system performs better at detecting helmets and non-helmet objects because it directs its attention to important spatial elements and benefits from AdamW's efficient weight regularization increases performance in industrial environments.

## Inference Time Comparison

Table 5 highlights the time to infer and processed speed of various object detecting models that were tested in this research. The results indicate that proposed enhanced hidden YOLOv8n architecture has the maximum real-time performance among the models discussed. Specifically, the improved model acquires 22.4ms/frame of inference (about 45 frames/second frame rate). Such a result implies that the system can handle video streams in real time without latency being a serious issue. The proposed model has a significantly faster processing rate compared to the previous models of detection like YOLOv3, which can execute at an approximate speed of 22/FPS and still has better detection rates.

**Table 5.** Real-Time Inference Performance Comparison

| **Model** | **Inference Time (ms/frame)** | **FPS** |
|---|---|---|
| YOLOv3 | 45.2 ms | 22 |
| YOLOv5 | 28.6 ms | 35 |
| YOLOv8n | 24.3 ms | 41 |
| YOLOv8n + CBAM | 26.1 ms | 38 |
| YOLOv8n + SSAM | 25.0 ms | 40 |
| Proposed Work | 22.4ms | 45 |

The inference rate of the lightweight detectors is 35 FPS and 41 FPS with YOLOv5s and YOLOv8n. The improved architecture increases the efficiency with the introduction of a Slim-Neck feature-fusion structure. The proposed architecture enables balancing the cost of computations and detection performance, the introduction of attention mechanisms like CBAM and SSAM slightly raises the computational cost. The GSConv-based Slim-Neck module decreases redundant feature computations and improves the efficiency of feature-aggregation provides faster inference than other variants that use the attention. The results of the conducted experiment support the assumption that the enhanced YOLOv8n model performs better in real-time processing, which makes it comfort application in industrial safety-monitoring systems where timely detection and reaction are the primary considerations

## 4.2 User Interface

Figure 7 shows the selected restricted zone by user. In live video feed, user can drag and draw the restricted zone, it is represented by green dashed rectangle and labeled as “Restricted Zone” this defines the active area for monitoring. Within restricted zone, the system will perform person and helmet detection to monitor activities. The controls on the left panel allow users to select or clear restricted regions and start the detection process when user clicks the “Start Detection” button. This demonstrates application and system allows flexible region selection for live detection monitoring in real-time safety detections. The user interface shows the real-time video feed where the enhanced YOLOv8 model has detected a person without wearing a helmet within the restricted zone, which is indicated by the red bounding box labeled “no helmet”. The Restricted zone is marked in blue color and the system status displays as LIVE, confirming active object detection with a frame rate of 44FPS.

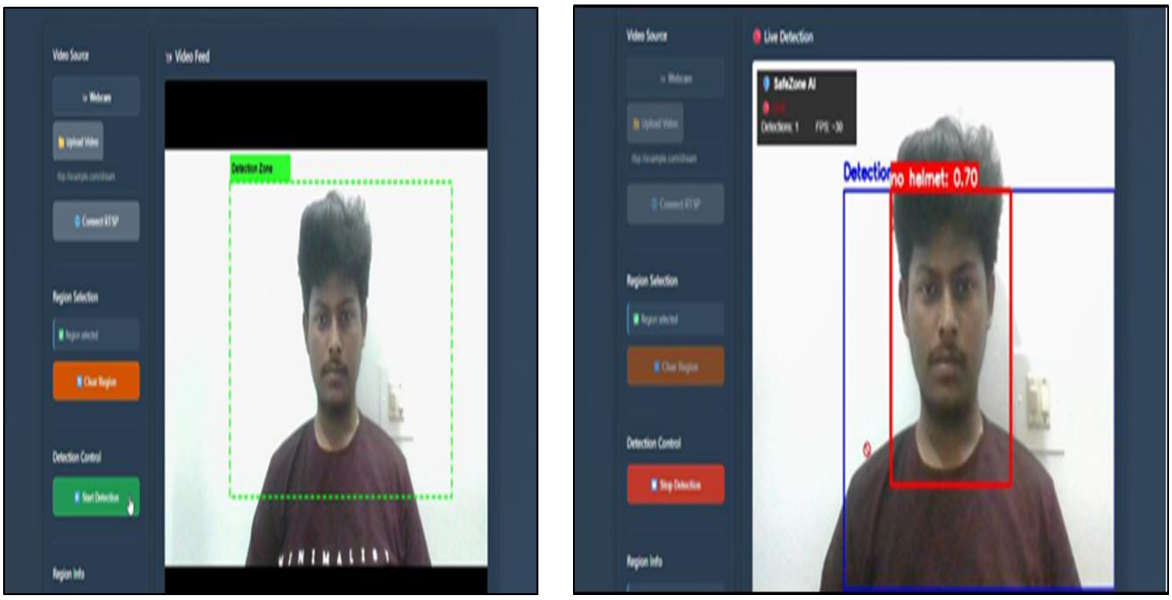


**Figure 7.** User Selects Restricted zone and Detection Results (No Helmet)

This alert system will send immediate alert message when a person is detected without proper safety equipment (helmet) within any restricted zone. This alert message is sent using the Twilio API. As the alert message shows, it provides clear and precise data about the type of violation, the number of people involved and the exact time of occurrence. This real-time alert mechanism improves workplace safety by bridging the gap between detection and response. This sort of proactive notification can play an important role in reducing accidents, improving compliance and sustaining safety levels across high-risk areas like construction sites and industrial areas.

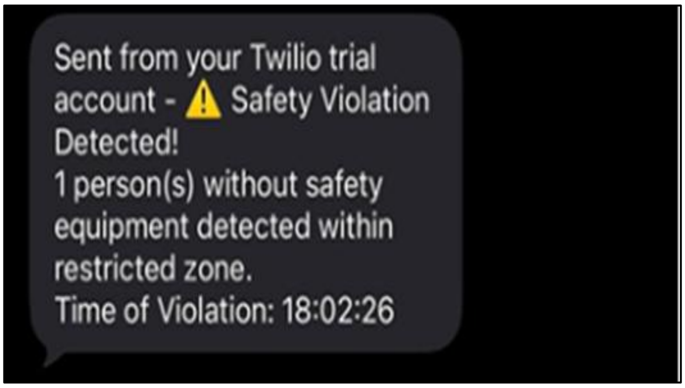


**Figure 8.** Alert Message After Violation

## 5. Conclusion

The Proposed real time safety helmet detection and restricted zone system has implemented successfully and integrated with enhanced YOLOv8 deep learning model, alert system to improve workplace safety in industries and constructions sites. Further improvements have been carried out on the YOLOv8 architecture by adding the Sparse Spatial Attention Mechanism (SSAM) and slim neck with GSConv and VoV-GSCSP modules, Optimization techniques like AdamW, Data Augmentation for better feature extraction and more precise in detection. Overall, this enhanced model obtained precision of 93.09%, recall of 91.63% and F1-score of 92.35% provide this model to be robust and reliable. Furthermore, the integration of a real-time alert system using Twilio API with web application interface developed using Angular and FastAPI, provides an effective end-to-end product for monitoring and alerting for safety violations. This system can automatically identify violations, monitor user marked restricted zones and send immediate alerts message via SMS. This will reduce manual supervision and human errors caused in industrial sites. Overall, this work presents a cost-effective, scalable and efficient safety monitoring system that can be used in different industries.

## References

[1] Lin, Bingyan. "Safety helmet detection based on improved YOLOv8." IEEE Access 12 (2024): 28260-28272.

[2] Jin, Peijian, Hang Li, Weilong Yan, and Jinrong Xu. "Yolo-esca: a high-performance safety helmet standard wearing behavior detection model based on improved yolov5." IEEE Access 12 (2024): 23854-23868.

[3] Liu, Yiping, Benchi Jiang, Huan He, Zhijun Chen, and Zhenfa Xu. "Helmet wearing detection algorithm based on improved YOLOv5." Scientific reports 14, no. 1 (2024): 8768.

[4] Zhao, Yu, Quan Chen, Wengang Cao, Jie Yang, Jian Xiong, and Guan Gui. "Deep learning for risk detection and trajectory tracking at construction sites." IEEE Access 7 (2019): 30905-30912.

[5] Han, Kun, and Xiangdong Zeng. "Deep learning-based workers safety helmet wearing detection on construction sites using multi-scale features." Ieee Access 10 (2021): 718-729.

[6] Tran, Van Than, Thanh Sang To, Tan-No Nguyen, and Thanh Danh Tran. "Safety helmet detection at construction sites using YOLOv5 and YOLOR." In International Conference on Intelligence of Things, pp. 339-347. Cham: Springer International Publishing, 2022.

[7] Zhao, Baining, Haijuan Lan, Zhewen Niu, Huiling Zhu, Tong Qian, and Wenhu Tang. "Detection and location of safety protective wear in power substation operation using wear-enhanced YOLOv3 algorithm." IEEE Access 9 (2021): 125540-125549.

[8] Liu, Ruihao, Zhongxi Shao, Zhenzhong Yu, and Rui Li. "Research on real-time helmet detection and deployment based on an improved YOLOv7 network with channel pruning." Signal, Image and Video Processing 19, no. 2 (2025): 118.

[9] Wang, Lijun, Yunyu Cao, Song Wang, Xiaona Song, Shenfeng Zhang, Jianyong Zhang, and Jinxing Niu. "Investigation into recognition algorithm of helmet violation based on YOLOv5-CBAM-DCN." Ieee Access 10 (2022): 60622-60632.

[10] Li, Longlong, Zhifeng Wang, and Tingting Zhang. "Gbh-yolov5: Ghost convolution with bottleneckcsp and tiny target prediction head incorporating yolov5 for pv panel defect detection." Electronics 12, no. 3 (2023): 561.

[11] Park, Man-Woo, Nehad Elsafty, and Zhenhua Zhu. "Hardhat-wearing detection for enhancing on-site safety of construction workers." Journal of Construction Engineering and Management 141, no. 9 (2015): 04015024.

[12] Mneymneh, Bahaa Eddine, Mohamad Abbas, and Hiam Khoury. "Automated hardhat detection for construction safety applications." Procedia engineering 196 (2017): 895-902.

[13] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587. 2014..

[14] Girshick, Ross. "Fast r-cnn." In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448. 2015.

[15] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015).

[16] Liu, Hai, Xiang Wang, Wei Zhang, Zhaoli Zhang, and You-Fu Li. "Infrared head pose estimation with multi-scales feature fusion on the IRHP database for human attention recognition." Neurocomputing 411 (2020): 510-520.

[17] Wang, Lili, Xinjie Zhang, and Hailu Yang. "Safety helmet wearing detection model based on improved YOLO-M." IEEE Access 11 (2023): 26247-26257.

[18] Li, Zhishan, Wenqing Xie, Lingzhi Zhang, Shan Lu, Lei Xie, Hongye Su, Weidong Du, and Weifeng Hou. "Toward efficient safety helmet detection based on YoloV5 with hierarchical positive sample selection and box density filtering." IEEE transactions on instrumentation and measurement 71 (2022): 1-14.

[19] Chen, Junhua, Sihao Deng, Ping Wang, Xueda Huang, and Yanfei Liu. "Lightweight helmet detection algorithm using an improved YOLOv4." Sensors 23, no. 3 (2023): 1256.

[20] Cheng, Rao, Xiaowei He, Zhonglong Zheng, and Zhentao Wang. "Multi-scale safety helmet detection based on SAS-YOLOv3-tiny." Applied Sciences 11, no. 8 (2021): 3652.