Cureus Journal of Computer Science

Open Access Original Article

Enhancing Occupational Safety: Real-Time Detection of Protective Equipment

Sahil Bhingarde [1] , Ketaki Chalke [1] , Niyati Kartar [1] , Mrunmayee Dalvi [1] , Jyoti More [1]

1. Computer Engineering, Fr. Conceicao Rodrigues Institute of Technology, Navi Mumbai, IND

Corresponding authors: Sahil Bhingarde, sahilbhingarde004@gmail.com, Ketaki Chalke, ketakichalke15@gmail.com, Niyati Kartar, niyatikartar@gmail.com, Mrunmayee Dalvi, mrunmayeedalvi08@gmail.com, Jyoti More, jyoti.more@fcrit.ac.in

Received 06/30/2025
Review began 08/17/2025
Review ended 02/13/2026
Published 02/21/2026

DOI:
https://doi.org/10.7759/s44389-026-00031-5

Abstract

In dangerous environments, such as building sites, maintaining worker safety is essential to reducing accidents and guaranteeing the adherence to safety rules. In order to monitor and enforce compliance with personal protective equipment (PPE) in real time, this study introduces an AI-based Safety Kit Detection system that makes use of computer vision and deep learning. The suggested solution uses YOLOv8, a sophisticated object detection model, to recognize vital safety equipment, including gloves, vests, and helmets. To determine whether PPE is present or not, the system analyzes live video streams, extracts frames, and uses deep learning techniques. It allows for prompt intervention by creating real-time notifications and updating a manager dashboard in the event of non-compliance. This kind of technology improves workplace safety, reduces human inspection errors, and ensures compliance with safety regulations by automating PPE monitoring. The high detection accuracy, robustness under various environmental situations, and real-time processing capacity are all demonstrated by the experimental findings. In the near term, scalability across sectors will be improved, predictive analytics will be incorporated, and detection will be made accessible to more types of PPE.

Categories: AI/ML-based decision support systems, Computer Vision, Deep Learning
Keywords: safety kit detection, ppe compliance, yolov8, computer vision, workplace safety, deep learning

Introduction

The correct and consistent use of Personal Protective Equipment (PPE) is essential for ensuring occupational safety across various high-risk sectors, including manufacturing, construction, and healthcare. Traditional methods for monitoring PPE compliance typically involve manual inspections, which are inherently labor-intensive and susceptible to human error. To address these limitations, recent research has focused on developing automated systems leveraging computer vision and deep learning techniques to enable real-time PPE detection and compliance monitoring.

One prominent approach utilizes the YOLO (You Only Look Once) object detection framework to identify safety equipment such as helmets, goggles, jackets, gloves, and protective footwear. A study employing the YOLOv7 model demonstrated notable accuracy, achieving a mean Average Precision (mAP) of 87.7 percent in detecting these items [1]. Building on these findings, subsequent research introduced an enhanced YOLOv8-based model incorporating advanced modules such as Dilated Windowed Residual, Atrous Spatial Pyramid Pooling, and Normalized Wasserstein Distance. This improved architecture achieved a mAP of 92.4 percent for helmet detection, underscoring its effectiveness in real-time safety applications [2].

Further studies have explored lightweight architectures aimed at real-time PPE detection in resource-constrained environments. For instance, a YOLOv4-tiny-based system was implemented in healthcare settings to monitor the donning and doffing processes of PPE. This system provided immediate feedback to medical personnel, ensuring adherence to proper procedures and minimizing the risk of cross-contamination [3].

The advancement of comprehensive and high-quality datasets has significantly contributed to the training and validation of these detection models. The SH17 dataset, comprising over 8,000 annotated images across 17 PPE classes, has played a pivotal role in enhancing detection performance. Models trained on SH17 have achieved accuracy levels exceeding 70.9 percent, demonstrating the dataset's practical utility in workplace safety systems [4]. Similarly, the PPED dataset, consisting of more than 15,000 images encompassing diverse PPE categories and environmental conditions, has facilitated robust model development and evaluation [5].

Recent innovations have also introduced multimodal fusion techniques, combining RGB and thermal imaging to improve detection reliability under challenging conditions, such as low illumination or occlusion. A study employing RGB-infrared image fusion demonstrated improved performance in detecting PPE in high-temperature and nighttime industrial environments [6].

How to cite this article
Bhingarde S, Chalke K, Kartar N, et al. (February 21, 2026) Enhancing Occupational Safety: Real-Time Detection of Protective Equipment. Cureus J Comput Sci 3 : es44389-026-00031-5. DOI https://doi.org/10.7759/s44389-026-00031-5

In addition, transformer-based architectures such as the Vision Transformer and the Swin Transformer have been investigated for PPE recognition tasks. By capturing long-range dependencies and complex spatial features, these models have exhibited superior performance compared to traditional convolutional neural network (CNN)-based approaches in visually complex and cluttered scenes [7].

System objectives and approach

The development of an effective PPE detection framework involves implementing the YOLOv8 model to identify essential safety gear such as helmets, vests, gloves, and boots. To enhance detection performance, the model is fine-tuned on a custom dataset comprising diverse workplace scenarios and varying environmental conditions. The system's performance is rigorously evaluated using standard object detection metrics, including mAP, precision, and recall. While precision and recall provide important insights into false positive and false negative predictions, respectively, they are sensitive to confidence thresholds and class imbalance when considered independently. As a result, they are not employed as standalone performance benchmarks. Instead, mAP is emphasized as the primary evaluation metric, as it provides a more comprehensive and balanced assessment by jointly accounting for classification accuracy and localization performance across all PPE categories. This evaluation strategy is aligned with safety-critical application requirements and established practices in existing literature, ensuring reliable and practical deployment in real-world industrial environments.

Building on the detection capabilities, a real-time safety monitoring system is developed to process live video streams. This system integrates an automated alert mechanism to notify supervisory personnel in the event of PPE non-compliance. To ensure efficient and low-latency performance in practical deployments, the system is deployed on edge computing hardware, such as the NVIDIA Jetson Nano, enabling real-time detection in dynamic industrial environments.

To maintain long-term system robustness and adaptability, performance optimization is conducted through periodic model updates. These updates incorporate newly collected data reflecting evolving PPE designs and changing workplace conditions. Furthermore, to align the system with regulatory expectations, collaboration with occupational safety and health authorities is pursued to ensure compliance with prevailing industry standards.

Despite notable progress, PPE detection systems continue to face persistent challenges. Detection accuracy may be adversely affected by occlusions, variations in PPE design, and environmental factors such as lighting, background clutter, and weather conditions. Future research should focus on the development of self-adaptive models capable of learning new PPE styles autonomously. In addition, the integration of depth sensors may improve detection in occluded scenarios, while further leveraging edge computing technologies will facilitate scalable, real-time deployment in diverse operational settings.

Literature survey

As industries look for scalable substitutes for manual safety inspections, automated detection of PPE has drawn more attention.

Early methods depended on traditional computer vision methods including rule-based segmentation and hand-crafted features, which had poor robustness against background clutter, illumination change, and occlusion. The development of deep learning-based object detection frameworks greatly increased the precision and dependability of detection in challenging industrial settings [1].

Due to their high speed-accuracy ratio, single-stage detectors as SSD and YOLO variations have become widely used to meet latency limitations [1,3,8,9]. CNN-based detectors have become popular among deep learning techniques. Although two-stage detectors like Faster R-CNN showed good localization accuracy for PPE detection, their high computational cost prevented them from being used for real-time monitoring [10]. Effective helmet and vest detection in construction and industrial environments was reported in several investigations using YOLOv3, YOLOv4, and YOLOv5; however, performance loss was frequently observed for small or partially occluded objects like gloves and goggles [3,9,10].

YOLOv7 and YOLOv8 architectures, which integrate better feature aggregation, decoupled heads, and improved optimization techniques, were recently developed [8]. While maintaining real-time inference capacity, these models showed significant improvements in mAP [2,11]. Through architectural changes, such as attention mechanisms, contrastive learning, and feature pyramid revisions, a number of studies improved YOLO-based PPE identification and achieved good results in controlled datasets [11]. Nevertheless, many of these studies focus on specific PPE categories (often helmets only), which limits their relevance to thorough safety compliance monitoring [2,8].

Another important factor in PPE detection studies has been the availability of datasets. Improved generalization was made possible by large-scale datasets like SH17, which allowed for multi-class PPE identification under various circumstances [4]. However, there are still unresolved issues with dataset imbalance, insufficient representation of incorrect PPE use, and a lack of annotated small objects [4,12]. Furthermore, deployment limitations and real-time viability on edge hardware are rarely discussed in research that only assess performance on carefully chosen datasets [5,9,13].

More recent efforts have explored multimodal sensing and transformer-based architectures to improve robustness under low visibility and cluttered scenes [6,7]. While these approaches yield promising accuracy improvements, they introduce increased computational overhead, making edge deployment more challenging [6].

Overall, existing literature demonstrates substantial progress in PPE detection accuracy; however, challenges persist in detecting small and visually similar PPE items, achieving reliable real-time performance on resource-constrained devices, and balancing dataset scale with annotation quality [5,9,11]. These limitations motivate lightweight, deployment-oriented detection frameworks accompanied by transparent evaluation and benchmarking.

Table *1* presents a quantitative comparison of representative PPE detection approaches reported in the literature. The comparison highlights differences in accuracy, runtime performance, dataset scale, and hardware configuration. Metrics are reported as stated in original works. Differences in datasets and class complexity are explicitly acknowledged.

Reference	Model	PPE Classes	Dataset Size	mAP@0.5 (%)	FPS/Latency	Hardware	Notes
[10]	Faster R-CNN	Helmet, Vest	~5k images	70–80	<10 FPS	GPU (Server)	High accuracy, not real-time
[9]	YOLOv5	Helmet, Vest, Gloves	~7k images	80–85	30–45 FPS	GPU	Performance drops for small PPE
[11]	YOLOv7 + Contrastive	12 PPE	~8k images	+2% over baseline	~25 FPS	GPU	Excludes gloves in some settings
[2]	YOLOv8 (helmet-only)	Helmet	~10k images	94.3	40+ FPS	GPU	Single-class detection
This work	YOLOv8m	6 PPE	319 images	86.2	~24 FPS	Low-end GPU	Multi-class, resource-efficient

TABLE 1: Quantitative benchmark comparison

R-CNN, Region-based Convolutional Neural Network; YOLO, You Only Look Once; PPE, Personal Protective Equipment; mAP, mean Average Precision; FPS, Frames Per Second; GPU, Graphics Processing Unit

Existing system

The contemporary landscape of PPE detection has evolved significantly with the integration of artificial intelligence and deep learning methodologies. Modern systems predominantly rely on object detection frameworks such as TensorFlow, PyTorch, and Keras, coupled with state-of-the-art models like YOLO for real-time identification of safety gear in industrial environments. These models are capable of detecting helmets, vests, gloves, and boots with varying degrees of accuracy and latency. Nevertheless, while each system offers unique advantages, they are often accompanied by limitations in scalability, detection robustness, or contextual adaptability, as summarized in Table *1*.

Research Gaps and Opportunities

Despite notable advancements in computer vision and object detection technologies, several research gaps remain unaddressed in existing PPE detection systems.

One persistent challenge is the detection of small objects, particularly at longer distances. Most object detection models exhibit reduced performance when identifying small-scale PPE, which can hinder effective monitoring on large construction sites or in surveillance footage with wide-angle views [12].

Additionally, data privacy concerns are becoming increasingly significant, particularly in contexts where continuous surveillance and personal data collection may infringe on workers' privacy. Ethical considerations regarding data handling, storage, and consent require further exploration and regulatory alignment [13].

Another underexplored area is the incorporation of predictive safety measures. Existing systems tend to focus solely on reactive monitoring, while predictive analytics - such as hazard anticipation or maintenance forecasting - could play a critical role in preemptively enhancing worker safety [14].

Furthermore, there is a notable lack of dataset diversity across existing research. Most datasets are constrained to specific environments or limited PPE categories. Expanding data collection to encompass varied industrial sectors and environmental conditions would contribute to developing more generalizable and context-aware detection models.

Problem Statement and Objectives

Identified challenges: Several technical challenges currently hinder the performance and reliability of PPE detection systems. Reflected surfaces and low-light environments often lead to distorted or missed detections due to inadequate image clarity and high contrast. Similarly, partially visible objects or improperly worn safety vests present difficulties in consistent identification, as key features may be obscured or not fully captured in the frame.

Another significant challenge is the visual similarity between certain objects, such as helmets and caps. These objects often share similar shapes and colors, making it difficult for traditional detection models to differentiate them accurately without advanced feature extraction and contextual understanding.

Problem statement: Ensuring on-site compliance with safety gear requirements remains a challenge due to the limitations of manual inspection processes. These procedures are not only time-consuming but also prone to oversight, particularly in large-scale operations. Additionally, many existing detection systems demonstrate limited accuracy, especially in cluttered or visually complex environments, reducing their effectiveness in critical safety applications. The lack of real-time and automated reporting capabilities further delays the identification of non-compliance, hindering timely corrective action.

Research objectives: This study aims to address the aforementioned limitations by developing a robust, scalable, and intelligent PPE detection framework. The primary goal is to design and implement a deep learning-based system capable of automatically identifying the presence or absence of essential safety gear such as helmets and gloves in real-time scenarios. In addition, the framework seeks to incorporate an integrated reporting module that can generate and disseminate safety compliance reports automatically via email or web-based dashboards, effectively highlighting any instances of missing or improperly worn PPE. Finally, the system is intended to efficiently analyze images and video data across diverse industrial environments, ensuring scalability for large datasets while maintaining high detection accuracy even in conditions involving occlusion, clutter, and variable lighting. A comparative summary of existing PPE detection systems, their methodologies, and identified research gaps is presented in Table *2*.

Attribute	Visionify PPE Detection [12]	Visual Detection of Personal Protective Equipment and Safety Gear on Industry Workers [13]	Procom AI Safety Detection [14]
Technology Stack	Uses TensorFlow or PyTorch with models like YOLO for real-time PPE detection, integrated with OpenCV for video processing and a modern frontend (likely React.js or Angular)	Employs AI models like YOLOv3/v4 for PPE detection, with cloud platforms such as AWS for scalable data processing and real-time monitoring	Utilizes deep learning frameworks like PyTorch or Keras along with YOLO for PPE detection, integrated with real-time alert systems (e.g., Twilio) and OpenCV for video analysis
Key Research Description	Developed and optimized YOLO models for PPE detection. Dataset preparation, model training and validation on Google Colab. Evaluation using precision, recall, F1 score, and mAP	Focused on detecting five PPE classes. Used Kaggle dataset and additional images from Google. Applied data augmentation techniques	Detects PPE use with AI-powered systems. Alerts for non-compliance to ensure continuous safety. Targets industrial and operational safety improvements
Gap Identified	Difficulty in detecting small objects at long distances. Limited to six PPE classes. Lack of diverse datasets from construction sites	Need for tracking methods and region-specific detection. Ethical concerns over privacy and security	Primarily targets general industrial applications, with less emphasis on integrating wearable technology or enhancing proactive safety measures

TABLE 2: Comparison of PPE detection systems

YOLO, You Only Look Once; PPE, Personal Protective Equipment; mAP, mean Average Precision

Materials And Methods

Introduction

This section describes the approach taken to create the Safety Kit Detection System using the YOLOv8 architecture. The methodology includes data collection, annotation, preprocessing, model training, and evaluation. The system's major goal is to use innovative algorithms based on computer vision to accurately identify five types of safety equipment: boots, goggles, gloves, vests, and helmets. The system workflow includes the steps mentioned below. The workflow diagram made by us here (Figure *1*) shows the entire process and highlights the framework's general structure by graphically illustrating the end-to-end pipeline, from data gathering and preprocessing to real-time detection, alarm production, and report preparation.

FIGURE 1: Workflow diagram illustrating the entire pipeline of the Safety Kit Detection System

Original diagram created by the author.

Data collection and preprocessing

The proposed system utilizes YOLOv8 (You Only Look Once, version 8), object detection model, to determine whether workers are equipped with the required PPE. The development process comprises multiple stages, beginning with data collection and preprocessing.

The initial phase involves the acquisition and preparation of a curated dataset, consisting of images and video frames of workers both wearing and not wearing safety gear. To improve robustness, the dataset includes variations in lighting conditions, camera viewpoints, and worker postures, as illustrated by representative samples.

Data cleaning procedures are applied to remove low-quality, blurred, or corrupted images. Image resolution and size are standardized to ensure uniform processing. To enhance variability and mitigate overfitting, data augmentation techniques are employed, including horizontal flipping, rotation, brightness adjustment, and saturation-based augmentation to simulate low-light conditions.

Although the manually annotated dataset comprises 319 original images, the effective training set exceeds 1,300 images when including augmented samples and neutral images representing non-compliant classes such as no helmet and no gloves. Data augmentation is applied exclusively to the training subset, while validation and testing are conducted using only the original, non-augmented images.

Image annotation and labeling are carried out using LabelImg [15], Roboflow [16], and CVAT [17]. Each safety item (e.g., helmet, vest, gloves, boots, goggles) is annotated with bounding boxes and labeled individually. Annotations are stored in the YOLO format using text files that specify bounding box coordinates.

Table *3* summarizes the class-wise distribution of images and annotated instances, demonstrating balanced representation across compliant, non-compliant, and neutral PPE categories. The dataset contains a total of 3,426 annotated instances across 395 images, including explicit non-compliance classes such as no helmet, no gloves, no boots, and no goggles, supporting reliable multi-class detection. While the dataset does not exhaustively cover all possible industrial environments, the class-level distribution provides sufficient diversity for pilot-scale validation.

Class Name	Images	Instances
person	395	395
helmet	266	266
vest	257	257
gloves	371	371
boots	290	290
goggles	131	131
no_helmet	266	266
no_gloves	371	371
no_boots	290	290
no_goggle	214	214
none	395	395
Total	395	3,426

TABLE 3: Class-wise distribution of images and annotated object instances across compliant and non-compliant PPE categories

PPE, Personal Protective Equipment

The curated dataset is split into three subsets: 70% for training, 20% for validation, and 10% for testing. This split supports effective model optimization and unbiased performance evaluation.

Prior to deblurring during preprocessing, each input frame is resized to a fixed spatial resolution of 640 × 640 pixels to standardize input dimensions and ensure consistent frequency-domain processing. No color space conversion or additional preprocessing operations are applied. The resized image is processed using Wiener filtering, where deblurring is performed by estimating the inverse of the degradation function while incorporating a noise-to-signal power ratio to regularize the solution. This approach mitigates noise amplification inherent in direct inverse filtering and produces stable image restoration. The preprocessing pipeline is illustrated in Figure *2*.

FIGURE 2: Pipeline for obtaining inference after preprocessing

Model performance is evaluated using key metrics such as mAP and the F1-score, which balances precision and recall. Techniques including threshold tuning and non-maximum suppression are employed to reduce false positives and enhance overall detection reliability.

Finally, the trained model is deployed in a real-time detection framework using live video feeds. Upon detecting absent or incorrectly worn safety gear, the system automatically generates alerts to notify supervisors, thereby supporting timely interventions and promoting workplace safety compliance.

Implementation of the system

Model Selection and Training

To develop an effective PPE detection system, multiple object detection architectures were evaluated, including Faster R-CNN, SSD, and variants of the YOLO model. Following a comparative performance analysis, the YOLOv8m model was selected due to its optimal trade-off between detection accuracy and inference speed, making it highly suitable for real-time applications. Additionally, the YOLOv8s model will be assessed to compare its efficiency and resource utilization relative to YOLOv8m.

Dataset Preparation

The dataset was curated to include several categories of safety equipment such as helmets, gloves, vests, boots, and goggles. A multi-step preprocessing pipeline was designed to ensure dataset quality and relevance.

Initially, images were collected from publicly available datasets and open-source platforms to guarantee diversity in environmental conditions, including varying lighting, camera angles, and backgrounds. Precise annotation of safety gear and instances of non-compliance was performed using the Computer Vision Annotation Tool, allowing for accurate bounding box placement.

During preprocessing, all images were resized to 640 × 640 pixels to standardize input dimensions. Pixel values were normalized to a (0,1) range. To eliminate low-quality data, the Laplacian Variance technique was applied to detect and discard blurred images. For those that could be recovered, Gaussian noise reduction and Wiener filtering were applied to improve image clarity. Additionally, the pHash method was employed to identify and remove duplicate images, reducing redundancy in the dataset.

To increase variability and improve model generalization, extensive data augmentation was carried out. This included image rotation, flipping, brightness and contrast adjustments, and color jittering-resulting in six variations per original image.

The dataset was then partitioned into training and validation subsets, allocating 80% for training and 20% for validation. An observed class imbalance in the “None” category, representing missing PPE, negatively affected initial model performance. To address this, additional samples were incorporated into this category to ensure balanced class representation across the dataset. The dataset curated by us is uploaded on Kaggle [18]

Model Training

Training was conducted using Google Colab, utilizing the PyTorch-based Ultralytics YOLOv8 framework. The system was configured with Python, CUDA, and PyTorch libraries, running on an NVIDIA Tesla T4 GPU to accelerate computations.

Pre-trained YOLOv8m weights were loaded to leverage transfer learning, thereby reducing training time and improving accuracy. The training configuration included a batch size of 16, image dimensions of 640 × 640 pixels, 70 training epochs, and the use of stochastic gradient descent with momentum as the optimization algorithm. A cosine decay schedule was adopted for learning rate adjustment throughout training.

The YOLOv8 training pipeline was used to initiate model training. Labels classified as “No Category,” which previously hindered performance improvements, were removed from the training set. Throughout training, key performance indicators such as training and validation loss, as well as precision and recall metrics, were monitored using Weights & Biases and TensorBoard for real-time visualization and tracking.

Model Optimization and Fine-Tuning

Upon completion of initial training, several optimization strategies were employed to further enhance the system’s performance. Hyperparameter tuning was performed using grid search to identify optimal values for learning rate, batch size, and weight decay.

To improve object localization accuracy, default anchor boxes were recalibrated to better match the dimensional distributions of the safety gear objects. Additionally, the non-maximum suppression threshold was fine-tuned to reduce the occurrence of false positives and maximize the model’s precision.

Deployment and Inference

The trained model is deployed in a near real-time detection framework using live 720p video feeds in a cloud-based environment. For 720p input frames, the system achieves approximately 11-12 frames per second (FPS) on a GT 710 GPU and approximately 24 FPS on an NVIDIA T4 GPU, with preprocessing constituting the primary computational bottleneck. Upon detecting absent or incorrectly worn safety gear, the system generates automated alerts to support timely safety interventions.

Future Enhancements

Several future improvements are planned to further increase the system’s accuracy and adaptability across diverse industrial settings. The dataset will be expanded to include a wider range of real-world PPE usage scenarios across various industries and environmental contexts.

To improve bounding box localization, advanced feature extraction and bounding box regression techniques will be explored. Furthermore, the adoption of federated learning is being considered to allow decentralized model updates without compromising data privacy, thereby facilitating scalable deployment across multiple industrial sites.

Ultimately, the final trained model will be optimized for large-scale industrial deployment, ensuring robust real-time safety monitoring and reliable compliance verification across dynamic workplace environments.

Results And Discussion

Results

Experimental Evaluation Overview

This section summarizes the experimental evaluation of the proposed PPE detection system, focusing on detection accuracy, real-time performance, and deployment behavior. Detailed descriptions of dataset preparation, preprocessing, and training configurations are provided in the Materials and Methods section.

The dataset used for evaluation includes multiple PPE categories such as helmets, gloves, vests, boots, and goggles, along with explicit non-compliance and neutral classes. All input frames were evaluated at a resolution of 720p and resized to 640 × 640 prior to inference, consistent with the experimental protocol described earlier.

The YOLOv8m architecture was evaluated based on its ability to balance detection accuracy and inference speed under constrained computational settings. Performance was assessed using standard object detection metrics, including precision, recall, mAP at IoU 0.50 (mAP@50), and mAP averaged across IoU thresholds from 0.50 to 0.95 (mAP@50:95).

A graphical user interface (GUI) was also evaluated as part of the system pipeline, enabling image and video-based inference and visualization of detection results. The GUI supports automated report generation in PDF format, providing structured summaries of PPE compliance for practical safety monitoring applications.

Real-Time Performance Evaluation

End-to-end runtime performance was evaluated using live 720p input frames resized to 640 × 640 prior to inference. On a system equipped with an Intel i5-10400F processor and a GT 710 GPU, the average preprocessing time per frame was approximately 160 ms, while model inference and post-processing required approximately 1.5 ms and 1 ms, respectively. This resulted in an effective throughput of approximately 11-12 FPS.

When deployed on an NVIDIA T4 GPU in a cloud-based environment, the system achieved approximately 24 FPS for the same input resolution. Across both hardware configurations, preprocessing constituted the dominant contributor to overall latency. All reported values reflect end-to-end measurements, including preprocessing, inference, and post-processing stages.

These results demonstrate that the proposed system achieves near real-time performance under low-resource GPU constraints and cloud-based deployment settings, supporting its applicability in practical industrial monitoring scenarios.

Model Performance and Evaluation

The YOLOv8m-based PPE detection system was evaluated using a dataset consisting of 319 original images containing 2,016 labeled object instances. Performance evaluation employed precision (P), recall (R), mAP@50, and mAP@50:95 as primary metrics.

The model achieved an overall precision of 0.860, recall of 0.830, and mAP@50 of 0.862, indicating reliable detection performance across multiple PPE categories. The mAP@50:95 value of 0.456 reflects moderate generalization across stricter localization thresholds.

Class-wise performance metrics are summarized in Table *4*, which reports detection results for individual PPE categories as well as the person and none classes. Helmet detection achieved the highest accuracy, while gloves and goggles exhibited comparatively lower precision and recall, likely due to their smaller size, frequent occlusion, and visual similarity to surrounding objects.

Class	Images	Instances	Precision (P)	Recall (R)	mAP@50
All	319	2016	0.86	0.83	0.862
Helmet	213	367	0.895	0.872	0.889
Gloves	121	245	0.801	0.731	0.752
Vest	218	347	0.876	0.836	0.878
Boots	132	307	0.797	0.736	0.817
Goggles	109	118	0.818	0.8	0.807
None	89	142	0.924	0.93	0.965
Person	296	490	0.906	0.904	0.925

TABLE 4: Performance metrics for safety kit detection

mAP, mean Average Precision

Comparative Analysis With Related Research

To contextualize the effectiveness of the proposed YOLOv8m-based PPE detection system, its performance was compared with reported results from widely used object detection models applied in safety monitoring tasks, including Faster R-CNN, SSD, and earlier YOLO variants.

Prior studies report that Faster R-CNN achieves mAP values typically ranging from 0.70 to 0.80 for PPE detection tasks; however, its higher computational complexity and slower inference speed limit its suitability for near real-time industrial monitoring. Reported results indicate that YOLOv5 offers a favorable trade-off between accuracy and speed, achieving mAP scores in the range of 0.80 to 0.85, though performance degradation has been observed for smaller objects such as gloves and goggles.

In comparison, the proposed YOLOv8m model achieved a precision of 0.860, recall of 0.830, and mAP@50 of 0.862, slightly outperforming YOLOv5 while maintaining competitive inference speed. Strong class-specific performance was observed for helmets (mAP@50 = 0.889) and vests (mAP@50 = 0.878), while detection of gloves and goggles remained more challenging due to intrinsic visual and spatial constraints.

Several recent studies report high detection accuracy under more restricted class settings. For example, a YOLOv8n-SLIM-CA model achieved an mAP@0.5 exceeding 94% for single-class helmet detection [2], and a Tiny YOLOv4-based approach reported low training loss when detecting five PPE categories [3]. A YOLOv7-based framework demonstrated modest mAP improvements across a 12-class PPE dataset [11], while other studies reported high accuracy for limited PPE combinations [9]. In contrast, the proposed system addresses a more complex multi-class detection problem that includes gloves, which are consistently reported as challenging to detect. Consequently, the achieved performance remains competitive and appropriate for the increased task complexity.

GUI and Report Generation

The system incorporates a user-friendly GUI designed to facilitate efficient PPE detection workflows. Users can easily upload images or videos through the interface, enabling seamless selection and processing of media files. This functionality is illustrated in Figure *2*, which demonstrates the intuitive design of the GUI that supports straightforward interaction for both image and video-based PPE detection tasks.

Once the image or video is up loaded, the system performs real-time detection and displays bounding boxes around detected PPE items, indicating whether workers are compliant or non-compliant, as illustrated in Figure *3*.

FIGURE 3: GUI for image/video upload: Users can upload media files for PPE detection

GUI, Graphical User Interface; PPE, Personal Protective Equipment

The image demonstrates the robustness of the model across a wide range of visual conditions, including close-up shots, outdoor scenes, zoomed-out perspectives, active construction sites, and varying levels of blur from Figure 4.1(ip) to Figure 4.5(ip). Across these diverse environments, the model is largely successful in detecting key safety-related classes such as helmets, gloves, vests, and boots, as clearly observed in Figure 4.1(op), Figure 4.2(op), Figure 4.3(op), and Figure 4.5(op). This indicates strong generalization and reliable performance in realistic scenarios. However, in Figure 4.4(op), a misclassification can be observed where a camera is incorrectly detected, which appears to stem from background coloration and visual clutter causing confusion for the model. Overall, the model handles all six defined classes correctly the majority of the time, with only occasional errors, highlighting its effectiveness while also pointing to limited edge cases that could benefit from further refinement.

FIGURE 4: Real-time detection: Bounding boxes around detected PPE items

PPE, Personal Protective Equipment

The system automatically generates a detailed PDF report summarizing PPE compliance findings. This report includes a pie chart illustrating the percentage of workers wearing each type of safety equipment, providing an at-a-glance overview of overall compliance rates. Additionally, a bar graph displays the number of unique PPE items worn by each individual, highlighting variations in safety kit completeness. The report also features a comprehensive table listing each detected person alongside any missing safety equipment, facilitating targeted follow-up actions. Visual representations included in the report are exemplified in Figure 5.

FIGURE 5: Report visualization: Pie chart showing PPE compliance, bar graph of unique equipment worn by each person, and table listing missing safety gear

PPE, Personal Protective Equipment

Discussion

The experimental findings show that the suggested YOLOv8m-based framework accomplishes accurate multi-class PPE identification across goggles, boots, helmets, vests, gloves, and non-compliance situations. Strong localization and classification performance is confirmed by the achieved mAP@50 of 86.2%, but a closer look exposes significant trade-offs pertinent to practical implementation.

Helmets and vests score the best in terms of detection accuracy among PPE categories. Because these objects have unique shapes and occupy greater spatial regions inside photos, this behavior is consistent with previous research. On the other hand, because of their tiny size, frequent occlusion, and visual closeness to background objects, gloves and goggles regularly provide lower precision and recall. This emphasizes a basic trade-off between detection reliability and thorough PPE coverage, particularly when going beyond coarse-grained categories.

When it comes to deployment, YOLOv8m strikes a balance between computing efficiency and precision. The suggested method allows near real-time inference on edge devices by sacrificing modest accuracy benefits in lieu of significantly lower latency as compared to heavier two-stage detectors. However, this efficiency comes at the expense of decreased robustness in low-resolution and extreme occlusion scenarios, indicating that temporal aggregation or hybrid sensing may further increase dependability.

Another crucial trade-off is the comparatively limited dataset size. The small scale limits generalization claims, even as careful curation and augmentation reduce overfitting and allow proof-of-concept validation. Therefore, rather than showing universal resilience, the results should be understood as proving feasibility. For large-scale deployment, increasing dataset diversity across industries, camera perspectives, and PPE designs is still crucial.

Finally, the integration of automated reporting and visualization introduces practical value beyond detection accuracy alone. By transforming raw detections into actionable compliance insights, the system supports operational decision-making. Nevertheless, this added functionality increases system complexity and emphasizes the importance of stable inference performance to avoid false alerts in safety-critical settings.

Overall, the proposed system occupies a pragmatic middle ground between accuracy, speed, and deployability. The findings reinforce that real-world PPE monitoring requires not only high detection metrics but also transparent evaluation, explicit limitations, and careful alignment between model complexity and deployment constraints.

Conclusions

This work presents an AI-driven PPE detection system based on the YOLOv8m architecture for automated safety compliance monitoring in industrial environments. The proposed system demonstrates reliable multi-class detection of personal protective equipment, including helmets, gloves, vests, boots, and goggles, achieving a mAP (mAP@50) of 86.2% with balanced precision and recall across categories. Experimental evaluation shows that the system operates in near real time (~24 FPS) under low-resource GPU constraints and cloud-based deployment settings, with preprocessing identified as the primary computational bottleneck.

While strong performance is observed for larger PPE items such as helmets and vests, detection of smaller objects, including gloves and goggles, remains comparatively challenging due to occlusion and visual similarity. In addition, the limited dataset size constrains generalization claims; therefore, the presented results should be interpreted as proof-of-concept validation rather than exhaustive industrial coverage. Beyond detection accuracy, the integration of GUI and automated report generation enables practical safety monitoring by transforming raw detections into actionable compliance insights. Future work will focus on expanding dataset diversity, improving small-object localization, and exploring advanced training paradigms such as federated learning to enhance scalability and adaptability across diverse industrial settings. Overall, the proposed system provides a balanced and deployable solution for PPE compliance monitoring, supporting safer workplaces through intelligent visual analytics.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Sahil Bhingarde, Niyati Kartar, Ketaki Chalke, Mrunmayee Dalvi, Jyoti More

Acquisition, analysis, or interpretation of data: Sahil Bhingarde, Niyati Kartar, Ketaki Chalke, Mrunmayee Dalvi, Jyoti More

Drafting of the manuscript: Sahil Bhingarde, Niyati Kartar, Ketaki Chalke, Mrunmayee Dalvi, Jyoti More

Critical review of the manuscript for important intellectual content: Sahil Bhingarde, Niyati Kartar, Ketaki Chalke, Mrunmayee Dalvi, Jyoti More

Supervision: Jyoti More

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Redmon J, Farhadi A: YOLOv3: An incremental improvement. arXiv. 2018, 10.48550/arXiv.1804.02767
2. Lin B: Safety helmet detection based on improved YOLOv8. IEEE Access. 2024, 12:28260-28272. 10.1109/access.2024.3368161
3. Verma A, Ghajari G, Jawad KMT, Salehi HP, Amsaad F: Real-time automated donning and doffing detection of PPE based on Yolov4-tiny. arXiv. 2024, 10.48550/arXiv.2407.17471
4. Ahmad HM, Rahimi A: SH17: A dataset for human safety and personal protective equipment detection in manufacturing industry. arXiv. 2024, 10.48550/arXiv.2407.04590
5. Ji X, Gong F, Yuan X, Wang N: A high-performance framework for personal protective equipment detection on the offshore drilling platform. Complex & Intelligent Systems. 2023, 9:5637-5652. 10.1007/s40747-023-01028-0
6. Chen YT, Shi J, Ye Z, Mertz C, Ramanan D, Kong S: Multimodal object detection via probabilistic ensembling. Computer Vision - ECCV 2022. ECCV 2022. Lecture Notes in Computer Science. Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T (ed): Springer, Cham; 2022. 13669:139-158. 10.1007/978-3-031-20077-9_9
7. Dosovitskiy A, Beyer L, Kolesnikov A, et al.: An image is worth 16×16 words: transformers for image recognition at scale. arXiv. 2021, 10.48550/arXiv.2010.11929
8. Elesawy A, Mohammed Abdelkader E, Osman H: A detailed comparative analysis of you only look once-based architectures for the detection of personal protective equipment on construction sites. Eng. 2024, 5:347-366. 10.3390/eng5010019
9. Lo J-H, Lin L-K, Hung C-C: Real-time personal protective equipment compliance detection based on deep learning algorithm. Sustainability. 2023, 15:391. 10.3390/su15010391
10. Nath ND, Behzadan AH, Paal SG: Deep learning for site safety: Real-time detection of personal protective equipment. Automation in Construction. 2020, 112:103085. 10.1016/j.autcon.2020.103085
11. Samma H, Al-Azani SH, Luqman H, Alfarraj M: Contrastive-based YOLOv7 for personal protective equipment detection. Neural Computing and Applications. 2023, 36:2445-2457. 10.1007/s00521-023-09212-6
12. Visionify's AI-powered PPE. https://visionify.ai/ppe-compliance.
13. Karlsson J, Strand F, Bigun J, Alonso-Fernandez F, Hernandez-Diaz K, Nilsson F: Visual detection of personal protective equipment and safety gear on industry workers. arXiv. 2022, 10.48550/arXiv.2212.04794
14. Procom Office Solutions, AI-Powered Safety Equipment Detection. https://www.procom.co.in.
15. Labelimg: Graphical image annotation tool. https://pypi.org/project/labelImg/.
16. Roboflow: AI-assisted annotation tool. https://roboflow.com/annotate.
17. CVAT: Computer vision annotation tool. https://www.cvat.ai/.
18. PPE Kit Detection (Construction Site Workers) dataset. https://www.kaggle.com/datasets/ketakichalke/ppe-kit-detection-construction-site-workers.