Real-Time PPE Detection and Safety Compliance Monitoring System for Construction Sites Using Machine Learning

Yashvardhan Rathore

M.Sc. Information Technology (Part 2)

D. G. Ruparel College of Arts, Science and Commerce, Mumbai, India

Abstract

Construction sites expose workers to falls, falling objects, impact, and visibility hazards, and Personal Protective Equipment (PPE) such as safety helmets, high-visibility jackets, and safety shoes is the most direct line of defence. Verifying PPE use is still mostly manual: supervisors walk the site or review CCTV footage, which is slow, inconsistent, and cannot cover every worker at all times. This paper proposes an AI-assisted system that automatically detects workers and their PPE and flags probable non-compliance in real time. A YOLO-based object detector is trained on an annotated multi-class dataset (Person, Helmet, Safety Jacket, Safety Shoes). Detections are passed to a rule-based compliance module that associates PPE items with individual workers and identifies missing equipment. The system accepts images, recorded video, and live camera streams through OpenCV, and presents worker counts, PPE status, confidence scores, and time-stamped violation evidence through a Streamlit dashboard. Performance is to be evaluated with Precision, Recall, F1-score, and mAP under varying lighting, camera angle, worker density, distance, and occlusion. The system is intended to support, not replace, human safety personnel. This paper presents the design, methodology, and anticipated outcomes of the proposed system; experimental results will be reported on completion of implementation.

Keywords: Personal Protective Equipment, PPE detection, YOLO, object detection, computer vision, deep learning, construction safety, real-time monitoring, safety compliance

1. Introduction

Construction is among the most hazardous industries. Workers operate simultaneously across different zones, performing tasks that carry risks of head injury, foot injury, and collisions with vehicles and machinery. PPE reduces these risks, but it only helps if it is actually worn. The common items on a construction site are safety helmets, high-visibility safety jackets, and safety shoes.

Traditional PPE enforcement relies on supervisors and safety officers observing workers directly or reviewing CCTV recordings. On large sites with many workers spread over wide areas, continuous manual monitoring is difficult and time-consuming. Safety personnel cannot observe every worker at once, and violations that occur out of sight go undetected.

Advances in deep learning and computer vision make automated monitoring practical. Object detectors can locate and classify multiple objects within a single frame at speeds suitable for video. The YOLO (You Only Look Once) family of detectors in particular predicts bounding boxes and class probabilities in a single network pass, which is well suited to real-time applications [3].

This work proposes a system that uses a YOLO-based detector to find workers and PPE items in images, recorded video, and live camera feeds, then applies compliance logic to determine whether each worker meets the required PPE standard. Missing PPE is treated as a potential safety violation, recorded with date, time, confidence score, and a captured frame, and displayed through an interactive interface for review by safety personnel.

1.1 Purpose of the Study

The primary aim is to develop an AI-assisted, real-time PPE detection and compliance monitoring system that identifies PPE usage among construction workers and helps safety personnel monitor compliance. The system is designed to support human decision-making by providing automated visual detection and a record of potential violations.

1.2 Problem Statement

Ensuring that every worker wears the required PPE throughout the working period is difficult with manual supervision alone. The main limitations are:

- Continuous monitoring of all workers is impractical.
- Large sites have workers distributed across many locations.
- Safety staff cannot observe every worker simultaneously.
- Manual review of CCTV footage is slow.
- PPE can be hard to identify because of distance, camera angle, lighting, or partial obstruction.
- Multiple workers often appear in the same frame and may overlap.

An intelligent vision-based system is therefore needed that can detect workers, helmets, jackets, and shoes; monitor several workers at once; identify missing PPE; process both recorded and live video; record violations; and be evaluated with standard machine-learning metrics.

2. Objectives of the Study

The project has the following objectives:

1. Collect and prepare construction-site images containing workers and PPE.
2. Preprocess and annotate the images for object-detection training.
3. Develop a YOLO-based object-detection model for PPE detection.
4. Detect construction workers in images and video frames.
5. Detect safety helmets worn by workers.
6. Detect safety or high-visibility jackets.
7. Detect safety shoes.
8. Develop a compliance mechanism that identifies missing required PPE.
9. Process recorded videos and live camera feeds for real-time monitoring.
10. Identify and record potential PPE safety violations.
11. Build an interactive interface for displaying detection and compliance results.
12. Study system performance under different site conditions: lighting, camera angle, worker density, distance, and partial occlusion.

3. Literature Review

3.1 YOLO-Based Object Detection

Redmon et al. introduced YOLO as a unified approach that treats detection as a single regression problem, predicting bounding boxes and class probabilities directly from full images in one evaluation [3]. This removed the multi-stage pipelines of earlier detectors and made high-speed detection feasible. Later versions improved accuracy and small-object performance, and modern implementations support images, video, and camera streams, making them a natural choice for real-time safety applications.

3.2 PPE Detection in Construction and Industrial Settings

Researchers have applied deep learning to detecting safety equipment on construction sites. Fang et al. developed a Faster R-CNN based method to detect non-hardhat use from far-field surveillance video [4]. Wu et al. presented a deep learning approach and benchmark dataset for detecting hardhats worn by construction personnel, including colour classification [5]. These studies show that learned detectors can identify workers with and without helmets across varied conditions.

Nath, Behzadan, and Paal extended the problem beyond helmets, using YOLO-based models to verify PPE compliance, including hardhats and vests, on construction sites. Their work used a dataset of 2,509 site images and incorporated alerting and time-stamped reporting for non-compliance [1]. More recently, Wang studied automated non-PPE detection using YOLOv10 and transformer architectures for surveillance and body-worn cameras, with benchmark datasets [2].

3.3 Challenges and Research Gap

Construction environments remain difficult for vision systems because of varying camera viewpoints, small or distant PPE, overlapping workers, cluttered backgrounds, changing illumination, and PPE hidden from the camera. Much existing work concentrates on helmets or helmets plus vests. Safety shoes are rarely addressed, partly because footwear is small, frequently occluded by terrain, materials, or other workers, and visually similar to ordinary shoes. This project therefore targets a multi-class detector covering helmets, jackets, and safety shoes, and links detections to individual workers for compliance monitoring.

4. Feasibility Study

Technical feasibility. The system can be built with open-source tools: Python, YOLO, PyTorch, OpenCV, Scikit-learn, and Streamlit. Training can run on a standard computer, with cloud GPUs available when needed.

Economic feasibility. Cost is minimal because the software stack is open source. Requirements are an existing laptop or desktop, an internet connection, and optional cloud GPU time.

Operational feasibility. The system works with images, recorded video, and live feeds. It highlights potential violations for human review before any safety action is taken.

Schedule feasibility. Work divides into manageable stages: dataset collection, dataset preparation, annotation, dashboard development, and testing and documentation.

Overall, the system is feasible with open-source technologies and available computing resources.

5. Research Methodology

The methodology follows twelve stages, summarised below.

5.1 Data Collection

Construction-site images with workers and PPE are gathered from suitable public datasets and, where appropriate, supplemented with additional images. The dataset should vary in worker position, number of workers, PPE types, and degree of occlusion.

5.2 Annotation

Images are annotated with bounding boxes using a tool such as Roboflow or CVAT. The initial classes are **Person, Helmet, Safety Jacket,** and **Safety Shoes**.

5.3 Preprocessing and Augmentation

The dataset is split into training, validation, and testing sets. Image preprocessing (such as resizing and normalisation) and augmentation (such as flipping, brightness and contrast changes, and scaling) are applied to increase diversity and robustness. The split should be done so that images from the same scene or video do not appear in more than one subset, to avoid inflated test scores.

5.4 Model Development and Training

A YOLO-based detector is trained on the prepared data. For every detected object the model outputs a bounding box, class label, and confidence score. Hyperparameters such as image size, batch size, number of epochs, and learning rate are tuned according to validation performance and available compute. Transfer learning from pretrained weights is expected to reduce training time and data requirements.

5.5 Model Evaluation

The trained model is evaluated on the held-out test set using Precision, Recall, F1-score, and mean Average Precision (mAP). Per-class results are reported, since small classes such as safety shoes are expected to perform differently from large classes such as Person.

5.6 PPE Compliance Analysis

After detection, a rule-based module analyses the output. PPE boxes are associated with a Person box using spatial relationships, for example a helmet near the upper region of the person box, a jacket overlapping the torso region, and shoes in the lower region. A worker is then classified as follows:

- Worker with Helmet + Jacket + Shoes: **PPE compliant**.
- Worker with Helmet + Jacket but no Shoes: **PPE violation (safety shoes missing)**.

The exact rules are configurable according to the PPE requirements chosen for a given site. Because absence of a detection is not proof of absence of PPE (the item may be occluded), the module should treat such cases as *potential* violations for human confirmation.

5.7 Image, Video, and Real-Time Detection

The model is tested on individual images, construction-site videos, and scenes with multiple workers, with results drawn as bounding boxes and labels. For live use, OpenCV captures frames from a webcam or compatible camera, and the pipeline is:

Camera → Video Frame → YOLO → Worker/PPE Detection → Compliance Analysis → Result

5.8 Violation Recording

When a worker is identified as potentially missing required PPE, the system records the date, time, violation type, confidence score, and the captured frame for later review by safety personnel. To avoid flooding the log with duplicates from consecutive frames, repeated alerts for the same violation within a short interval can be suppressed.

5.9 Application Development

A Streamlit interface provides image upload, video upload, live camera detection, PPE detection results, and worker counts.

5.10 Final Evaluation

The completed system is tested under varied conditions, including lighting, camera angle, worker density, distance, and occlusion, and evaluated on detection accuracy, Precision, Recall, and F1-score.

5.11 Facilities Required

Hardware: minimum Intel Core i5 or AMD Ryzen 5, 8 GB RAM, adequate storage; recommended 16 GB RAM and a dedicated GPU where available; a webcam or CCTV/video source; optional cloud GPU for training.

Software: Python 3.x, PyCharm, YOLO, PyTorch, OpenCV, NumPy, Streamlit, on Windows or Linux.

6. Discussion

No experiments have yet been run for this proposal, so this section discusses the expected behaviour and design considerations rather than reported results.

Expected detection performance. Based on prior work [1], [2], [4], [5], helmet and jacket detection are expected to perform well because these items are large, distinctive, and often brightly coloured. Safety shoes are likely to be the weakest class because of their small size, frequent occlusion, and similarity to everyday footwear. Reporting per-class metrics will make such differences visible rather than hidden within a single average.

Value of compliance logic. A detector alone only says what is in the frame. The compliance module turns detections into a decision about each worker, which is what safety personnel actually need. Its correctness depends on both detection quality and the association rules, so errors in either propagate to the final result. Evaluation should therefore cover worker-level compliance decisions, not only object-level mAP.

Role of human oversight. The system is designed to flag *potential* violations with supporting evidence (frame, timestamp, confidence). Human review before action reduces the consequences of false positives and false negatives and keeps accountability with safety staff.

Effect of site conditions. Distance, camera angle, lighting, crowding, and occlusion directly affect detectability. Testing across these conditions, as planned, should show where the system is dependable and where camera placement or additional training data is needed.

Practical trade-offs. Larger YOLO variants generally give higher accuracy but lower frame rates. The final model choice should balance accuracy against the speed needed for smooth real-time operation on the available hardware.

7. Expected Outcomes

The project is expected to deliver:

1. Automated detection of workers in images and video.
2. Detection of helmets, safety jackets, and safety shoes, including several PPE items in one scene.
3. Worker-level PPE compliance determination and recording of potential violations.
4. Real-time monitoring from live camera or video feeds.
5. Quantitative evaluation with Precision, Recall, F1-score, and mAP.
6. An interactive dashboard showing detected workers, PPE, compliance status, and violations.
7. An AI-assisted approach to supporting PPE monitoring at construction sites.

8. Limitations

- **Dataset dependence.** Performance depends on the quality, size, and diversity of the annotated data. A model trained on public datasets may generalise poorly to a particular site's uniforms, equipment colours, or camera setup.
- **Small and occluded objects.** Safety shoes and distant PPE are hard to detect reliably, and a missed detection can be mistaken for missing PPE.
- **Visual limits of PPE verification.** The system can detect that a helmet-like object is present but cannot confirm it is fastened correctly, certified, or in good condition. Similarly, it cannot tell certified safety shoes from ordinary shoes by appearance alone in many cases.
- **Rule-based association.** Spatial matching of PPE to workers can fail in crowded scenes where workers overlap.
- **Environmental factors.** Poor lighting, rain, dust, glare, and unfavourable camera angles reduce accuracy.
- **Hardware constraints.** Real-time speed on a CPU-only machine may be limited, restricting model size or frame rate.
- **Scope.** The system covers three PPE classes. Other PPE (gloves, goggles, harnesses) and hazards such as unsafe behaviour or restricted-zone entry are out of scope.
- **Privacy and ethics.** Continuous video monitoring of workers raises privacy and consent concerns; deployment should follow applicable regulations and workplace policies.

9. Future Scope

- Extending detection to gloves, safety goggles, ear protection, face masks, and fall-protection harnesses.
- Adding multi-object tracking so each worker is followed across frames, giving more stable decisions and fewer duplicate alerts.
- Evaluating newer detector architectures, including transformer-based and newer YOLO versions, as in recent work [2].
- Deploying on edge devices (such as embedded GPUs) for on-site, low-latency inference.
- Integrating alerts via SMS, email, or messaging apps, and storing violations in a database for trend analysis and reporting.

- Combining PPE detection with zone-based rules, for example requiring certain PPE only in designated hazardous areas.
- Adding worker identification, with appropriate privacy safeguards, for safety record-keeping.
- Using pose estimation to improve association of PPE to the correct body part.
- Domain adaptation and continual learning to adapt the model to new sites with limited new data.

10. Conclusion

This paper proposed a real-time PPE detection and safety compliance monitoring system for construction sites based on YOLO object detection, rule-based compliance analysis, and a Streamlit dashboard. By covering helmets, safety jackets, and safety shoes, and by linking detections to individual workers, the system aims to give safety personnel a practical, low-cost aid for continuous monitoring. Implementation and experimental evaluation are the next steps, and the findings will determine how well the approach holds up under real site conditions.

References

1. N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep learning for site safety: Real-time detection of personal protective equipment," *Automation in Construction*, vol. 112, p. 103085, 2020.
2. S. Wang, "Automated non-PPE detection on construction sites using YOLOv10 and transformer architectures for surveillance and body worn cameras with benchmark datasets," *Scientific Reports*, vol. 15, art. 27043, 2025.
3. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, real-time object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.
4. Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T. M. Rose, and W. An, "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," *Automation in Construction*, vol. 85, pp. 1–9, 2018.
5. J. Wu, N. Cai, W. Chen, H. Wang, and G. Wang, "Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset," *Automation in Construction*, vol. 106, p. 102894, 2019.