International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 11s, 2025
https://theaspd.com/index.php

PPE Patrol: YOLO To The Rescue! – Advanced Object Detection For Enhanced Safety Compliance

Prof. Ajaj Khan[1], Prof. Jyotsana Goyal[2], Prof. Nisha Bhati[3], Prof. Kumar Gaurav[4], Aditya Kochhar[5], Aabhash Rathore[6], Abhijit Singh Mandloi[7]
[1] Assistant Professor Computer Science and Engineering Medicaps University Indore, India
Ajaj.khan@medicaps.ac.in
[2]Assistant Professor Computer Science and Engineering Medicaps University Indore, India
jyotsana.goyal@medicaps.ac.in
[3]Assistant Professor Computer Science and Engineering Medicaps University Indore, India
nisha.bhati@medicaps.ac.in
[4]Assistant Professor Computer Science and Engineering Medicaps University Indore, India
gaurav.kumar@medicaps.ac.in
[5]Student Computer Science and Engineering Medicaps University Indore, India
adityakochhar00@gmail.com
[6]Student Computer Science and Engineering Medicaps University Indore, India
aabhahshrathore1234@gmail.com
[7]Student Computer Science and Engineering Medicaps University Indore, India
mandloi.abhijit@gmail.com

***Abstract*–** *In spite of various safety precautions, the construction sector still encounters more fatalities in comparison to the other industries. While Personal Protective Equipment (PPE) is designed to prevent/mitigate any accidents, workers often overlook its use, whether unintentionally or on purpose. Manually monitoring safety compliance is challenging due to the large workforce, though ensuring their safety remains a top priority for site managers. Performing manual checking of safety mechanisms is troublesome because of a large number of workers available on-site but maintaining their security is topmost in site authorities' considerations. We conceived an automatic PPE recognition system based on computer vision (CV) technology to tackle this problem. The system is capable of recognizing various categories of PPE. As a part of this study [1], we also proposed a new dataset named CHVG, containing four colored hardhats, vests, safety glasses, person bodies, and person heads, with a total of eight distinct classes. Approximately 2100 images constitute the dataset that we have utilized, and each of them containing details of these eight classes labeled on it. We employed the use of the YOLOv8, which is well known for being anchor-free in its architecture, for the detection process. YOLOv8 is faster and more accurate than existing object detection models. The possibility of employing cutting-edge computer vision techniques to enhance safety monitoring activities in hazardous settings, such as construction sites, is demonstrated by this study.*
***Keywords*-** *Personal Protective Equipment (PPE), YOLOv8, Object Detection, Construction Safety, Occupational Hazards, Safety Compliance, Automated Monitoring, Real-Time Detection, CHVG Dataset.*

Ⅰ. INTRODUCTION

According to comprehensive data published by the International Labor Organization (ILO), every year the global workforce faces approximately 270 million workplace accidents and around 160 million occupational health issues. Alarmingly, these unfortunate events result in nearly two million fatalities annually. This figure accounts for roughly 4% of the total yearly deaths worldwide. Furthermore, about 15% of the global population is directly or indirectly affected by such incidents. The situation becomes increasingly severe in developing countries where regulatory frameworks are often weak, and enforcement mechanisms are inadequate. For instance, Turkey recorded the highest number of fatal work accidents among European Union countries in 2017, with its fatal accident rate per hundred thousand employees being 4.5 times greater than the EU average. In India, the problem is equally concerning. Construction workers in India represent only 7.5% of the world's labor force, yet they endure 16.4% of the total global occupational hazards. Research shows that accidents are more prevalent on smaller construction sites, where 62.8% of workers have reported incidents, compared to

47.4% on larger sites. According to these figures, occupational health and safety regulations are not only inadequate but also vary greatly depending on the size and location of the industry. Cutting-edge technology like the YOLO V8 detection system provides intriguing answers to these problems.

Ⅱ. LITERATURE REVIEW

Deep learning is poised to enhance automation in construction management by making it simpler with gadgets such as YOLO V8, which achieves real-time object detection with minimum error. With YOLO V8, tracking compliance to PPE usage becomes effortless for companies which in turn can reduce the accidents and risks associated. Aside from taking safety measures into account with automation, this technology also fosters a culture of responsibility and rational behavior on the part of workers. The relevance of Personal Protective Equipment (PPE) kits as a component of safety protocols is vital, particularly in situations that entail extensive and often daily contact with dangerous substances. PPE kits are designed with the sole intent of offering protection from chemicals and biological, as well as physical risks to the worker's wellbeing and safety. Application of PPE kits in the health sector has been made with particular emphasis to public health emergencies, which require a high rate of disease spread control coupled with the observance of necessary safety procedures and protocols. The adherence to strong compliance monitoring has been enhanced with the addition of technology, including YOLO V8, in the detection systems of PPE kits. The adoption of digital health technologies points to the need of supervising protocol compliance in real time, as indicated in the recent works of Ferdous.M et al.[3]. Furthermore, the empirical evidence points towards the relationship between appropriate PPE usage and improved workplace safety outcomes, indicating that regular compliance can significantly minimize workplace hazards [4]. Now let's take a look at YOLO V8's technical setup, and how it works to identify PPE compliance, and how it might transform safety procedures in risky industries in the section that follows.

Ⅲ. WORKING OF YOLO

Real-time object detection progressed significantly in 2016, when Joseph Redmon and others introduced the YOLO (You Only Look Once) process. YOLO processes region proposal, feature extraction, and classification all in one single, fast, efficient neural network. This differs from the conventional object detection methods that decompose the process into multiple steps. This cannot be accomplished with a single step in cardiac X-ray pictures, in that the single method identifies the objects and their respective bounding boxes in one single step, which increases speed and efficiency of the overall process compared to conventional methods. YOLO has experienced evolutionary changes since it was invented [6], making it more productive in terms of usability, accuracy, and speed. Thus, YOLO is more precise and efficient than the methods of the past [5]. Following is the structure of the YOLO model framework as indicated in Figure 1. The following steps are:

1. Input Image: The process starts by feeding an image into the system.
2. Grid Formation: Each cell in the grid created by segmenting the image is in charge of identifying items that fall within its field of view.
3. Feature Extraction: A convolutional neural network (CNN) is passed through each cell to extract features. The CNN has been trained on a large collection of images so that it can detect features which can help in detecting objects
4. Objectness Score: An objectness score is assigned to each cell to determine the likelihood of an object existing within it, calculated using logistic regression techniques.
5. Class Probability: The system predicts both the object class and the likelihood of that class for each cell, with conditional probabilities calculated using a SoftMax function.
6. Bounding Box: For each cell that predicts the presence of an object, YOLO calculates a bounding box that tightly surrounds the detected object. This box is defined by its center coordinates (x, y), along with its width and height.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 11s, 2025
https://theaspd.com/index.php

7. Redundant Detection Removal: Non-Maximum Suppression (NMS) is applied to discard overlapping or less confident bounding boxes, improving prediction precision.
8. Final Output: The model delivers a collection of bounding boxes, each tagged with its class type and confidence score.

Figure 1-Structure of YOLO.

IV. COMPARATIVE STUDY

Now we will compare the YOLO versions from v1 to v8 and check their efficiency for a better understanding.

Comparative Study of YOLO Versions:

YOLO Version	Methodology	Dataset	Image Size	Precision/Recall
YOLO v1	The input image is divided into a 7x7 grid in the first edition. If the center of each grid cell is inside that cell, then that cell will predict the items.	Trained on the PASCAL VOC and ImageNet datasets.	448x448	accuracy of 86% and a mean Average Precision (mAP) of 54%.
YOLO v2	Enhanced through the application of a neural network to simultaneously forecast bounding boxes and class probabilities across the entire image.	Trained on public drone datasets and YouTube video data.	416x416	88.25% precision and 85.44% recall.
YOLOv3	Introduced bounding box clustering using k-means and used four-dimensional feature maps for more accurate predictions.	Trained on ImageNet and COCO datasets.	416x416 320x320	84.14% precision and 89.27% recall.
YOLOv4	Incorporated advancements such as CSPDarknet, Spatial Pyramid Pooling, and Path Aggregated Networks to enhance performance.	Trained on the COCO dataset.	416x416 512x512	Achieved an impressive 89.32% precision and 92.48% recall.
YOLOv5	Known for its versatility, YOLOv5 uses CSPNetBackbone4, SPP, and advanced feature extraction techniques.	It's trained on the Kaggle dataset.	224x224	An impressive precision rate of 94.7%.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 11s, 2025
https://theaspd.com/index.php

YOLOv6	Anchor-free paradigm, SimOTA Tag Assignment Policy, and SIoU bounding box regression loss [7].	Trained on COCO datasets.	640x640	Achieved 43.3% Average Precision.
YOLOv7	Streamlines object recognition in real time using compound and extended scaling methods.	Trained on Kaggle and COCO datasets.	640x640	Achieved 56.8% Average Precision.
YOLOv8	The latest version integrates advanced data augmentation strategies, improving generalization.	The COCO dataset.	640x640	Precision of 96.5% and a recall rate of 87%..

Table 1 – Comparative study of YOLO models

V . SYSTEM TRAINING

For the project, we custom-trained the YOLOv8 model on the COCO dataset. On completion of the training period for the model which was evaluated using multiple performance metrics, we got these outputs such as:

1. Performance Metrics

- On testing the mean Average Precision (mAP) the output was 90.5%, which indicated high detection accuracy.
- On testing for the precision of the proportion of correctly identified PPE instances among all image detections was 92%.
- The recall proportion of the PPE instances that were successfully detected was 95%.

The model accurately detects whether a person wore the PPE kit or not as per the safety regulations, we can see the output in Figure 2.

Figure 2- Training Dataset

2. Confusion Matrix Evaluation

In Figure 3, the confusion matrix was generated to assess any misclassifications. The results demonstrated that there were both minimal false positives and false negatives, leading to the confirmation of the model's reliability in distinguishing and identifying different PPE elements.

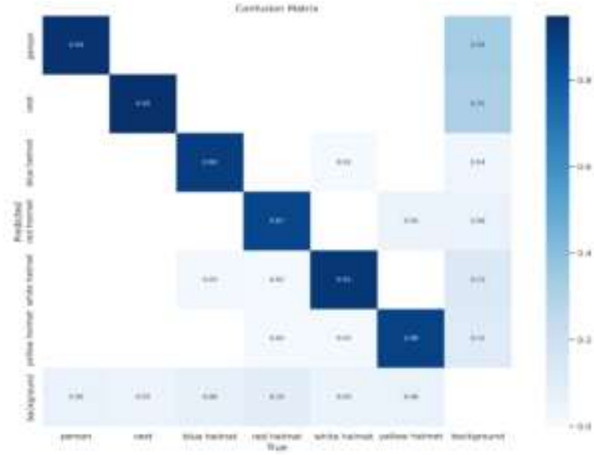

Figure 3- Confusion Matrix of YOLOv8.

3. Convergence and Loss Trends

Training and validation loss values stabilized after approximately 50 epochs, indicating efficient learning. The loss numbers kept going down oversome time, which showed that we managed to avoid over-fitting by using tricks like adding more varied training examples and tweaking the model's settings to keep it in check.

4. Real-Time Inference Speed

The YOLOv8 model processed the images in a period of 7.2 milliseconds, making it suitable and ideal for real-time applications in industrial and construction settings. In Figure 4 we can see the performance outputs of the model.

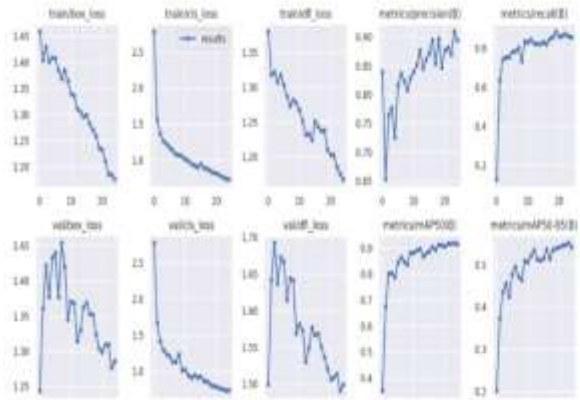

Figure 4-Performance Outputs of YOLO v8 model.

VI. RESULTS

After training, implementation of the model was the next step, the model was then prepared for real-world scenarios this was done by converting it into multiple formats to enhance compatibility across various hardware platforms:

- For deployment on edge devices and embedded systems, the ONNX format is optimized.
- Applications built using PyTorch can use the TorchScript format.
- The TFLite format was created for lightweight and portable applications.

1. Edge Device Deployment

The trained model was tested on an NVIDIA Jetson Nano, where it successfully detected PPE in real-time, maintaining a frame rate of ~30 FPS. In Figure 5 we can see that the subjects were not wearing any

PPE kits and the modeldisplayed no helmet, no vest, etc. on respective body parts as the output and takes a screenshot.

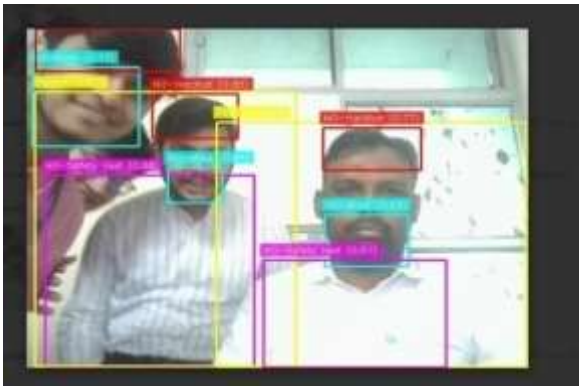

Figure 5- Real-time detection.

2. Integration with Surveillance Systems

The model was connected to a laptop and could also be connected to the CCTV surveillance systems, enabling automated monitoring of PPE compliance at construction sites. The system was programmed to:

- Trigger alerts if a worker is detected without the required PPE.
- Log compliance reports for safety audits.
- Improve workplace safety monitoring by reducing manual intervention.
- Granting permission to enter the construction site.
- Optimized to send a mail to the manager for each person allowed and not allowed on the site

As we can see in Figure 6, the model detects a person but not the equipment/PPE kit. So the model has been programmed as such to send a mail to the registered manager/person in a burst time period of every 10 seconds as each person passes through the camera. The model can only detect approximately 3-5 people at once accurately and if more people are present in the camera-frame, the model will throw errors in detection.

Figure 6- Email Sent.

VII. CONCLUSION

In developing countries like India, the construction sector experiences a high rate of fatalities, largely due to inadequate worker education and limited awareness of safety protocols [2]. In order to overcome this difficulty, the current study introduces an automated approach for identifying PPE infractions on-site using YOLOv8. The new CHVG dataset [1] was successfully used to train the model, allowing for precise detection of human anatomy features as well as protective gear like glasses, safety vests, and helmetsThe strategy also addressed the problem of under-reporting [2] by creating compliance with safety practices through training and observer contacts in line with worldwide trends to minimize risks in the workplace. Implementing the compliance strategy provided a sound framework within which personal/professional reporting practices were established and could be captured. Further studies may allow optimization of the system by utilizing multi-modal sensor data, or expanding the dataset to include factors influencing various settings, in order to support flexibility and rigour in diverse contexts. These findings indeed point to the possibility of AI-enabled computer vision radically redoing safety practices around the world in high-risk industries. The findings of this study identified that all models exhibited similarly high precision and recall for PPE identification, with YOLOv8x and YOLOv8l performing the best and exhibiting impressive accuracy and robustness [8].

Ⅷ. FUTURE SCOPE

In order to improve the relevance and dependability of our research and development initiatives, we also intend to collect real-time data from ongoing building sites [8]. This will yield crucial practical insights and by doing that we will be able to improve our methodology with this approach, making it more reliable and based on actual situations.The overarching goal is to design and translate a comprehensive system that leverages YOLO-based PPE detection to ensure automatic detection assignment of safety features. To facilitate the PPE data evaluation at the planning stage and assist in the production of an intelligent safety knowledge model that includes considerations for PPE usage, this system will also integrate a monitoring system to accumulate the important data. The framework also seeks to evaluate the degree of compliance, provide real-time monitoring capabilities with prompt remedial actions to address any possible hazards, and create fast notifications in the event of violations.

This model improves on earlier iterations by increasing inference speed with the addition of YOLOv10. By eliminating the non-max-suppression phase, which customarily eliminated superfluous bounding boxes, it achieves this. Instead, YOLOv10 uses a novel dual assignment strategy, incorporating both one-to-one and one-to-many assignment heads that are jointly optimized during training but rely on a single head during inference for faster processing. With this innovation and a 46% decrease in trainable parameters, the performance is almost identical to that of YOLOv9-c, indicating impressive efficiency [9].Looking forward, the expansion of PPE detection systems to include a broader range of safety gear — such as gloves, specialized boots, and industry-specific PPE — will significantly improve the adaptability and scope of these technologies [10].The future of YOLO-based PPE detection is highly promising, with industries increasingly moving toward smart safety solutions. Real-time detection systems will soon be integrated with IoT-enabled cameras to automatically issue alerts and generate compliance reports. Enhanced models will improve accuracy in complex environments, while cloud-based dashboards will enable safety trend monitoring across multiple sites. Furthermore, combining PPE detection with behavioral analysis and predictive risk modeling will transform workplace safety monitoring from a reactive measure into a proactive system, aligning with the global ESG and safety standard.

REFERENCES

[1] E. Guney, H. Altin, A. Esra Asci, O. U. Bayilmis, and C. Bayilmis, " YOLO-Based Personal Protective Equipment Monitoring System for Workplace Safety", jitsi, vol. 5, no. 2, pp. 77 - 85, Jun. 2024.

[2] Sasmita Samanta, Jyotiranjan Gochhayat, Critique on occupational safety and health in construction sector: An Indian perspective, Materials Today: Proceedings,Volume 80, Part 3, 2023, Pages 3016-3021, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.05.707. (https://www.sciencedirect.com/science/article/pii/S2214785321049701).

[3] FERDOUS M, AHSAN SMM. 2022. PPE DETECTOR: A YOLO-BASED ARCHITECTURE TO DETECT PERSONAL PROTECTIVE EQUIPMENT (PPE) FOR CONSTRUCTION SITES. PEERJ COMPUTER SCIENCE 8:E999 HTTPS://DOI.ORG/10.7717/PEERJ-CS.999.

[4] Gajić, Boško, Gajić, Katarina, Kaitović, Željko, Kresović, et al., "Effects of Irrigation Rate and Planting Density on Maize Yield and Water Use Efficiency in the Temperate Climate of Serbia", East Sarajevo: Faculty of Agriculture, 2023, https://core.ac.uk/download/595948802.pdf (accessed: 12 Feb, 2025).

[5] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[6] www.labellerr.com › blog › evolution-of-yolo-objectEvolution of YOLO Object Detection Model From V5 to V8

[7] A Comparative Study of Various Versions of YOLO Algorithm to Detect Drones Gayathridevi K1 , Dr. S. Kanmani2 Department of Information Technology (Puducherry Technological University), Puducherry, India E-mail: 1gayathridevi.k@pec.edu, 2kanmani@pec.edu

[8] Barlybayev, A., Amangeldy, N., Kurmetbek, B., Krak, I., Razakhova, B., Tursynova, N., & Turebayeva, R. (2024). Personal protective equipment detection using YOLOv8 architecture on object detection benchmark datasets: a comparative study. Cogent Engineering, 11(1). https://doi.org/10.1080/23311916.2024.2333209

[9] Hafiz Mughees Ahmad, Afshin Rahimi,SH17: A dataset for human safety and personal protective equipment detection in manufacturing industry, Journal of Safety Science and Resilience, 2024, ISSN 2666-4496, https://doi.org/10.1016/j.jnlssr.2024.09.002.

[10] Likhar, Shreyas, Vivek Tripathi, and Ketki Kshirsagar."Detecting Personal Protective Equipment Using YOLO."Journal of Emerging Technologies and Innovative Research (JETIR), vol. 11, no. 4, Apr. 2024, www.jetir.org.

[11] Rishi Agarwal, Sathwik Gundala, G S S Chalapathi. "Hardware-Based Implementation of Target Tracking in Unmanned Aerial Vehicles (UAVs)", 2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM), 2023

[12] Ergasheva, A.; Akhmedov, F.; Abdusalomov, A.; Kim, W. Advancing Maritime Safety: Early Detection of Ship Fires Through Computer Vision, Deep Learning Approaches, and Histogram Equalization Techniques. Fire 2024, 7, 84. [Google Scholar] [CrossRef]

[13] Zhang, Z.; Tan, L.; Tiong, R.L.K. Ship-Fire net: An improved YOLOv8 algorithm for ship fire detection. Sensors 2024, 24, 727. [Google Scholar] [CrossRef]

[14] Kim, D.; Ruy, W. CNN-based fire detection method on autonomous ships using composite channels composed of RGB and IR data. Int. J. Nav. Arch. Ocean Eng. 2022, 14, 100489. [Google Scholar] [CrossRef]

[15] Lee, H.G.; Pham, T.N.; Nguyen, V.H.; Kwon, K.R.; Lee, J.H.; Huh, J.H. Image-based Outlet Fire Causing Classification using CNN-based Deep Learning Models. IEEE Access 2024, 12, 135104–135116. [Google Scholar] [CrossRef]

[16] Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In European Conference on Computer Vision; Springer International Publishing: Cham, Germany, 2020; pp. 213–229. [Google Scholar]

[17] Tan, M.; Le, Q.E. Rethinking model scaling for convolutional neural networks. arXiv 2019, arXiv:1905.11946. [Google Scholar]

[18] Cheknane, M.; Bendouma, T.; Boudouh, S.S. Advancing fire detection: Two-stage deep learning with hybrid feature extraction using faster R-CNN approach. Signal Image Video Process. 2024, 18, 5503–5510. [Google Scholar] [CrossRef]

[19] Han, H. A novel single shot-multibox detector based on multiple Gaussian mixture model for urban fire smoke detection. Comput. Sci. Inf. Syst. **2023**, 20, 32. **[Google Scholar]** **[CrossRef]**