Seeing Safety: Computer Vision for Real-Time PPE Monitoring in the Middle East Construction Sector

Syed Suhail Ahmed[1], Drishty Sobnath[1], Cristina Turcanu[1], Ikram Rehman[2]

[1]Heriot-Watt University Dubai
[2]University of West London

syedsuhail21.ahmed@gmail.com, d.sobnath@hw.ac.uk, cristina.turcanu@hw.ac.uk, ikram.rehman@uwl.ac.uk

Abstract

Workplace safety is still a major concern worldwide, especially in high-risk settings like construction sites where manual PPE monitoring is insufficient. As employers want to move towards a more efficient and safer worksite, context awareness is becoming increasingly important in the construction industry. Besides, conventional PPE compliance checks can be ineffective, costly and prone to errors. This paper presents a novel approach to enhance workplace safety in construction environments through the development of an automated Personal Protective Equipment (PPE) detection system utilizing advanced computer vision techniques. YOLOv5 object detection models were used in this study to address the limitations of manual PPE compliance monitoring by achieving real-time detection under diverse environmental conditions, such as variable lighting and occlusion. The methodology encompasses comprehensive dataset preparation, annotation, and model training, achieving a mean Average Precision (mAP) exceeding 70%, with YOLOv5m(a) attaining an mAP@0.5 of 0.841. This research contributes to reducing workplace hazards by improving monitoring efficiency and lays a foundation for future advancements in industrial safety systems.

Introduction

The physical infrastructure that serves modern societies is shaped by the construction sector, which is one of the core pillars of economic growth in the Middle East region. However, the construction industry is also one of the most hazardous sectors worldwide, with significant challenges in ensuring worker safety on job sites. Evidence suggests that the rates of occupational accidents and fatalities in the Middle Eastern Crescent significantly exceed those observed in more established market economies (Awwad, El Souki and Jabbour 2016). The research involved creating and evaluating a deep learning (DL) computer vision model, 'Seeing Safety', which will identify PPE compliance in real time.

Numerous studies have applied YOLO-based detectors for PPE monitoring, often training separate models for different tasks or applying complex hybrid optimizations. For instance, (Alateeq, P. P. and Ali 2023) trained distinct YOLOv5 models for equipment and worker detection, while (Hussein Samma et al. 2023) implemented a modified YOLOv7 with contrastive loss to improve classification of visually similar PPE. We also incorporate automatic label correction and update using YOLOv9 (Ahmad and Rahimi 2024), reducing manual annotation errors—a step absent in earlier work.

Methodology

This section discusses the process for designing and creating a real-time PPE detection system for construction safety.

Step 1: Input Image: Captures unprocessed data from a construction site, showing workers with PPE.
Step 2: Preprocessing the Image: Resizes the image to 640x640 pixels to meet YOLO input requirements.
Step 3: Feature Extraction using YOLO: Uses the CSPDarknet backbone to extract features from the pre-process image.
Step 4: Multi Scale Detection: Employs the PANet neck with evolved anchors for multi-scale feature fusion, generating three-scale predictions to detect PPE items.
Step 5: Post Process for No-PPE Detection: Identify missing PPE by comparing detections against expected PPE items, dynamically drawing corresponding bounding boxes. If the item is present and overlaps the region, no box is shown
Step 6: Aggregate Results and Log Data: Logs detection and results.
Step 7: Output Image: Presents the final image with bounding boxes and labels.

Three metrics, precision, mean average precision and recall were used to evaluated how quickly and accurately the models performed.

Implementation

Dataset

The Pictor-PPE (Nath, Behzadan and Paal, 2020), VOC2028 (njvisionpower 2023), and CHV (Wang et al.,

2021) datasets were cleaned and combined into a single dataset of 9,695 images as part of the pre-processing process. Their characteristics are summarized in Table 1.

Table 1: Dataset Details

Dataset	Images	Class
Pictor-PPE	784	Person, Helmet, Vest
VOC20208	7581	Person, Helmet
CHV	1330	Person, Vest, Yellow[H], White[H], Red[H], Blue[H]

Exploratory data analysis was used to learn more about the makeup of the dataset and spot any possible issues with model training. As shown in Table 2, the class distribution has a notable class imbalance, with 20,525 person instances, 7,997 helmet instances, and only 1,155 vest instances.

Table 2: Class Distribution

Class	Instance Count
Person	20,525
Helmet	7997
Vest	1,115

The YOLOv5m(a) (Jocher, 2020) training started with label generation using YOLOv9. This improved label accuracy for small objects by detecting 17 classes. This was later reduced to six classes as shown in Table 4. Since the default anchors were not ideal, anchor evolution was then carried out for 100 generations, optimizing anchor boxes. Mosaic and mix-up augmentations were applied to increase resilience such as lighting and occlusion. Augmentations such as rotation and shear increased robustness to varying angles and perspectives in construction environments, while mosaic improved contextual understanding of PPE.

The configurations of these models, YOLOv5m(a), YOLOv5s(a), and YOLOv5s(b) are summarized in Table 3.

Table 3: Model Configurations

Model	Epochs	Batch	Freeze	Resolution	Dataset
YOLOv5s(a)	100 / -	8	0 / -	640x640	Original
YOLOv5s(b)	50 /-	6	10 / -	640x640	Original
YOLOv5m(a)	30 / 100	4	8 / 0	896x896	Extended

Evaluation

The YOLOv5m model is assessed for real-time PPE detection in this section.

Table 4 shows YOLOv5m(a) achieves the highest overall with mAP@0.5=0.841, excelling in "mask" with mAP@0.5=0.969 and "helmet" where mAP@0.5=0.901. Gloves" recall remains low (R=0.555), indicating persistent small object challenges. The FPS of 25-30 supports real-time use.

Table 4: Models' Performance Metrics of YOLOv5m(a)

Class	Instances	P	R	mAP@0.5
All	21772	0.864	0.776	0.841
Person	15319	0.860	0.816	0.890
Vest	881	0.852	0.762	0.818
Helmet	3484	0.873	0.833	0.901
Mask	348	0.948	0.945	0.969
Gloves	880	0.790	0.555	0.655
Glasses	810	0.860	0.744	0.812

The confusion matrix in Figure 1 shows that the model demonstrates strong performance for most classes, with high true positive rates along the diagonal. The 'Person' class achieves a true positive rate of 0.85, indicating actual instances are correctly identified, though 15% are misclassified as background. The 'Vest' and 'Helmet' classes also exhibit robust performances, with true positive rates of 0.78 and 0.87, respectively.

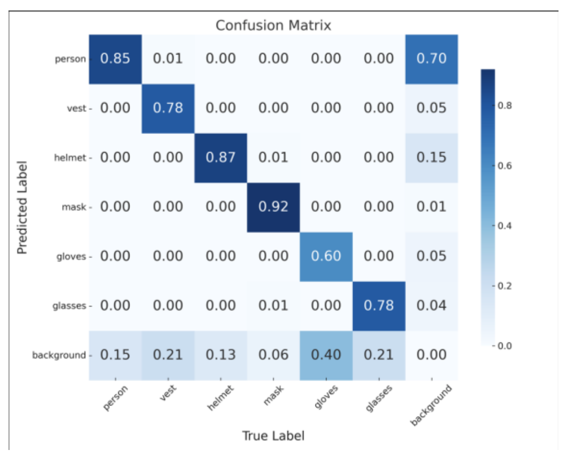

Figure 1: Confusion Matrix of YOLOv5m(a)

Conclusion

This project successfully developed YOLO models for real-time detection of PPE equipment to enhance safety compliance in the construction industry. YOLOv5m(a) model achieved notable results delivering a high mAP@0.5 of 0.841. A key innovation in the "Seeing Safety" project was the real-time No-PPE detection, achieved without training an additional class, by dynamically drawing bounding boxes to identify missing PPE. A larger dataset could boost model accuracy and robustness, making it a more competitive option for real-time applications in the Middle East construction sector.

Acknowledgments

This work was supported by the School of Mathematical and Computer Sciences Department of Heriot-Watt Dubai.

References

Ahmad, H.M., and Rahimi, A. 2024. SH17: A Dataset for Human Safety and Personal Protective Equipment Detection in Manufacturing Industry. *Journal of Safety Science and Resilience*. https://doi.org/10.1016/j.jnlssr.2024.09.002.

Alateeq, M.M., P. P., F.R. and Ali, M.A.S. (2023). Construction Site Hazards Identification Using Deep Learning and Computer Vision. *Sustainability*, [online] 15(3), p.2358. doi:https://doi.org/10.3390/su15032358.

Awwad, R.; El Souki, O.; and Jabbour, M. 2016. Construction Safety Practices and Challenges in a Middle Eastern Developing Country. *Safety Science*, 83: 1–11. https://doi.org/10.1016/j.ssci.2015.10.016.

Hussein Samma; Al-Azani, S.; Luqman, H.; and Alfarraj, M. 2023. Contrastive-Based YOLOv7 for Personal Protective Equipment Detection. *Neural Computing and Applications*, 36(5): 2445–2457. https://doi.org/10.1007/s00521-023-09212-6.

Jocher, G. 2020 Ultralytics/YOLOv5, GitHub. Available at: https://github.com/ultralytics/yolov5. Accessed: 2025-03-21.

Nath, N.D.; Behzadan, A.H.; and Paal, S.G. 2020. Deep Learning for Site Safety: Real-Time Detection of Personal Protective Equipment. *Automation in Construction*, 112: 103085. https://doi.org/10.1016/j.autcon.2020.103085.

njvisionpower. 2023. SafetyHelmetWearing-Dataset (安全帽佩戴检测数据集). GitHub. https://github.com/njvisionpower/Safety-Helmet-Wearing-Dataset. Accessed: 2025-04-25

Wang, Z.; et al. 2021. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors*, 21(10): 3478. https://doi.org/10.3390/s21103478.