

A Computer Vision-Based System for Detecting Safety Helmet Compliance on Construction Sites Using YOLOv5s

Minerva A. Paz Bodero¹ , Piero E. Suarez Chavez² 

¹t1022700821@unitru.edu.pe, t1512700521@unitru.edu.pe

^{1,2}Escuela Profesional de Ingeniería Informática, Universidad Nacional de Trujillo, Perú

Received: August 16, 2025 / Accepted: September 28, 2025 / Published: 5 June, 2026

doi: <https://doi.org/10.26439/ciii2025.8657>

ABSTRACT—The use of safety helmets is a critical measure for protecting construction workers; however, noncompliance remains a recurrent and high-risk issue. This paper presents a real-time computer vision system for helmet detection based on the YOLOv5s algorithm. The model was trained on more than 7 000 annotated images and deployed through a lightweight, scalable pipeline. Experimental results achieved a mean Average Precision (mAP at 0.5) of 91.9% and an optimal F1-score of 0.89 at a confidence threshold of 0.41, with an inference speed of 110 FPS. These findings demonstrate the system's effectiveness under real-world conditions, providing accurate and fast detection suitable for on-site safety monitoring and contributing to improved compliance in construction environments.

Index Terms—Computer vision, convolutional neural networks, deep learning, object detection, safety systems.

I. INTRODUCTION

According to 2020 data from the U.S. construction industry, the sector accounted for 21.2% of all occupational fatalities (1008 deaths) while representing only 4.1% of the total workforce, indicating that construction workers were 5.57 times more likely to die on the job than workers in other sectors [1]. In the same year, head injuries accounted for nearly 6% of nonfatal injuries resulting in days away from work [2]. These statistics highlight a critical safety gap, underscoring the importance of proper head protection—such as helmet use—in reducing the risk of severe injury.

According to 2021 U.S. data, the construction sector accounted for 25% of all occupational head injuries, making it the leading contributor to work-related traumatic brain injuries (TBIs) [3]. Falls on construction sites represented 68% of TBI cases, with falling objects contributing an additional 12% [3]. This underscores that safety helmets are not merely PPE but a critical barrier against severe injuries in high-risk occupational settings.

Several computer vision approaches have been proposed to address this issue. Hayat and Morgado-Dias [4] applied YOLOv5x to helmet detection, achieving a mean Average Precision (mAP) of 92.44% at 45 FPS, demonstrating robust performance even under low-light environments. Qian and Wang [5] introduced SHDet, a lightweight YOLOv5-based variant optimized through reverse attention and inverted residual blocks, which achieved 92.2% mAP with an inference time of only 3 ms per image, particularly suitable for embedded systems. An *et al.* [6] further improved You Only Look Once, version 5—small variant (YOLOv5s) by incorporating global attention, CBAM modules, and SIoU loss, achieving superior accuracy and speed compared to YOLOv3–v6. Li *et al.* [7] used convolutional neural networks in engineering contexts, reporting nearly 90% precision and over 85% recall, however, real-time deployment was not evaluated.

Previous research underscores the promise of deep learning techniques to enhance occupational safety; however, challenges remain regarding scalability, inference efficiency, and practical integration.

In response to this gap, this study proposes a real-time computer vision system for helmet compliance monitoring on construction sites, implemented using YOLOv5s. The main contribution is the demonstration of competitive detection accuracy with significantly higher inference speed on accessible hardware, enabling scalable deployment and supporting the enhancement of safety culture in construction environments.

II. METHODOLOGY

A. YOLOv5s Algorithm

Multiple studies indicate that YOLOv5s has evolved significantly from earlier versions, distinguishing itself through high inference speed and strong accuracy in real-time object detection [8].

Introduced by Ultralytics in 2020, YOLOv5 employs an architecture based on CSPDarknet (Cross Stage Partial, an enhanced Darknet backbone) and incorporates advanced techniques such as Mosaic Augmentation—a data augmentation strategy that combines four training images into one, improving recognition across different scales and object positions. The main steps of the YOLO detection process are summarized in the flowchart shown in Fig. 1, which illustrates the pipeline from image scaling through bounding-box prediction, confidence thresholding, and final output after non-maximum suppression.

B. Dataset

Preliminary adjustments were conducted on a small set of 45 images; however, the final training and validation were performed using an expanded dataset of 7,542 annotated images. This larger dataset ensured reliable detection of helmet use across diverse construction site conditions and scenarios.

This approach enables evaluation of the system’s validity and effectiveness under real-world situations, ensuring a high level of accuracy and reliability in safety helmet detection.

The dataset used in this study was obtained from the Roboflow platform [9], a widely used tool for dataset management and preprocessing in computer vision applications.

In this case, the YOLOv5-Hard Hat Detection Computer Vision Project was used, which includes two dataset versions comprising a total of 7,542 images depicting individuals with and without construction helmets.

These images were annotated by the dataset creator with two classes labeled *hat* and *person*.

For system development, the second version of the Roboflow dataset was selected, as shown in Fig. 2; it contains annotated images of individuals with and without safety helmets.

Examples of the labeling process are illustrated in Fig. 3, where bounding boxes identify workers in construction scenes. The dataset defines two original classes, *hat* and *person*, as illustrated in Figs. 4 and 5). For this study, these classes were relabeled as *protected* (helmet worn) and *at risk* (no helmet) to align with the project’s objectives.

C. System Design

Fig. 6 illustrates the overall workflow of the system, encompassing dataset preparation, model training, and real-time deployment.

The pipeline begins with the acquisition of training images, which undergo preprocessing steps such as normalization and resizing to enhance YOLOv5 model learning. The processed dataset is then used to train the network, enabling reliable identification of helmet use across diverse construction-site conditions.

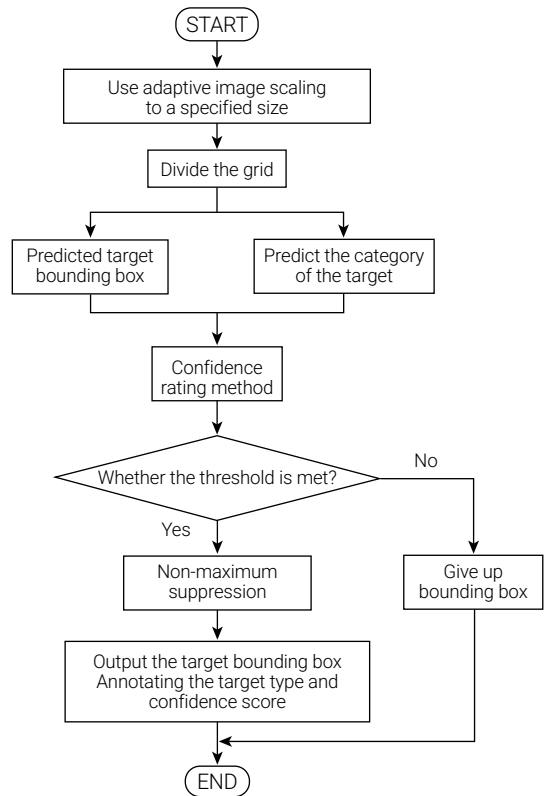


Fig. 1. Flowchart of the YOLOv5 algorithm.

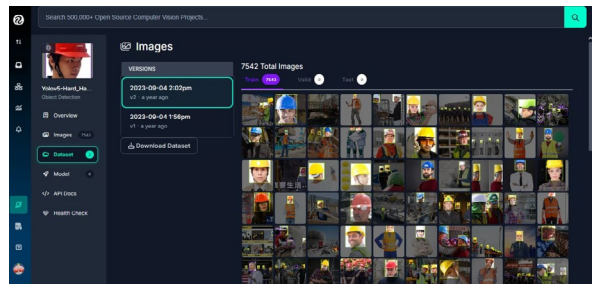


Fig. 2. Dataset used from Roboflow platform.

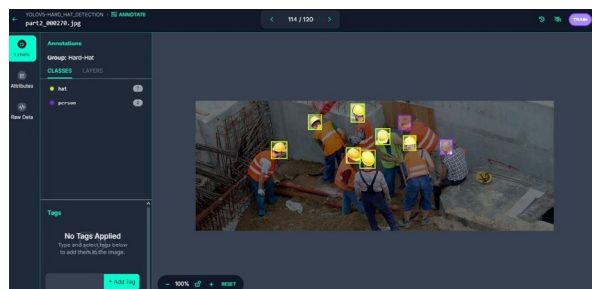


Fig. 3. Visualization of a labeled image.



Fig. 4. Visualization of the hat class.



Fig. 5. Visualization of the person class.

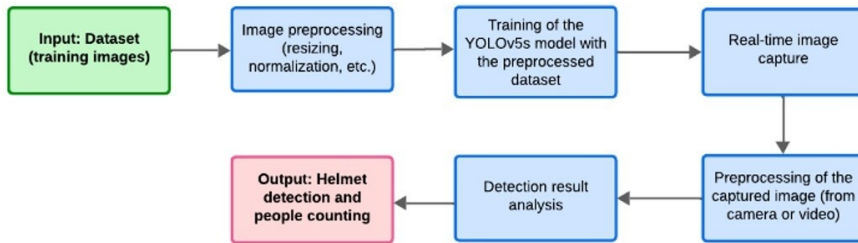


Fig. 6. Block diagram of the proposed system.

TABLE I
DISTRIBUTION OF IMAGES IN THE DATASET

Subset	Number of images
Training	4,568
Validation	2,296
Testing	678
Total	7,542

TABLE II
TRAINING CONFIGURATION PARAMETERS

Subset	Number of images
Input resolution	640 x 640
Batch size	16
Number of epochs	50
Initial weights	yolov5s.pt
Configuration file	dataset.yaml

After the training phase concludes, the workflow advances to real-time deployment. As shown in Fig. 7, images acquired from cameras or video streams undergo the same normalization steps before being analyzed by the trained model.

The system then interprets the detections, distinguishing between individuals with proper helmet use and those without protection, while also enabling personnel counting. In this way, the system not only identifies compliance but also highlights potential risks. The final stage generates outputs that support safety monitoring by providing accurate and timely information on helmet usage at construction sites.

D. Training Setup

The dataset used in this study was obtained from the Roboflow platform (YOLOv5 Hard Hat Detection v2), containing annotated images of individuals with and without safety helmets.

For the purposes of this research, the classes were relabeled as *protected* (helmet worn) and *at risk* (no helmet). Table I summarizes the number of images assigned to training, validation, and testing.

To enhance generalization, preprocessing included resizing all images to 640×640 pixels, normalization, and data augmentation (horizontal flips and brightness and contrast adjustments).

The system was implemented using the YOLOv5s model, which is well recognized for its efficiency in real-time object detection. Training was conducted in Google Colab, while deployment and testing were performed in Visual Studio Code. The experimental environment consisted of a laptop equipped with an AMD Ryzen 7 7435HS processor, 24 GB of RAM, an NVIDIA GeForce RTX 4060 GPU, and a 512 GB solid-state drive.

Table II details the key training parameters used in this work.

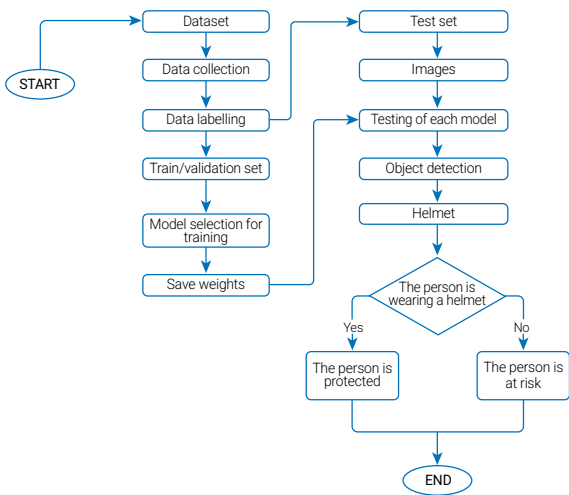


Fig. 7. Flowchart of the proposed system.

E. Metrics

The effectiveness of the proposed system was evaluated using standard object-detection metrics. Precision quantified the proportion of correct positive detections, whereas recall indicated the model's ability to identify all relevant instances. Their harmonic mean, expressed as the F1-score, served as a comprehensive metric balancing both aspects of detection quality.

A confusion matrix was also employed to visualize prediction outcomes, distinguishing correct classifications from recurrent errors and thereby helping to identify misclassification trends between the *protected* and *at risk* categories.

Furthermore, the precision-recall curve was analyzed to examine how sensitivity and specificity varied across different decision thresholds, offering additional insight into model robustness. Finally, training and validation statistics were tracked across epochs to verify stable convergence and to assess the model's generalization performance on unseen samples.

F. Scope and Limitations

The proposed system demonstrates strong performance in differentiating between individuals who comply with helmet regulations and those who do not, even under diverse environmental and lighting conditions. Its flexibility supports deployment in both indoor and outdoor construction areas, and the architecture is scalable, enabling integration with multiple cameras across large sites to achieve comprehensive coverage.

Moreover, the model can be retrained and incrementally updated with new data, enhancing its adaptability to different contexts and improving long-term robustness.

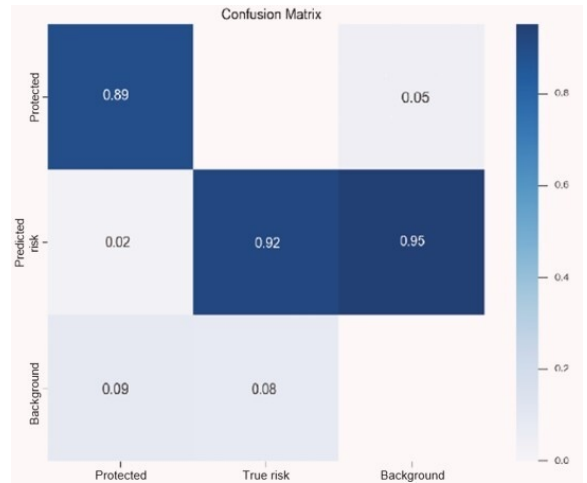


Fig. 8 Confusion matrix.

Despite these advantages, some challenges remain. False positives and false negatives may still occur, particularly when helmets are partially occluded or when visually similar objects are present.

Although YOLOv5s is considered lightweight compared to larger architectures, real-time deployment still requires adequate GPU resources, which may increase hardware costs. In addition, system accuracy may degrade under extreme lighting conditions—such as heavy shadows or excessive brightness—highlighting the need for further robustness improvements in uncontrolled environments.

III. RESULTS

A. Confusion Matrix

As shown in Fig. 8, the confusion matrix evaluates the safety-helmet detection model across three classes: *protected*, *at risk*, and *background*.

For the *protected* class, the model correctly identifies 89% of instances, with 2% being misclassified as *at risk* and 9% as *background*. The *at risk* class achieves the highest accuracy, with 92% of cases correctly detected, whereas 2% are misclassified as *protected* and 8% as *background*.

Finally, the *background* class exhibits greater confusion, with 95% accuracy; however, 5% of samples are misclassified as *protected* or *at risk*.

Overall, the confusion matrix indicates strong performance in distinguishing between the critical safety-related categories (*protected* and *at risk*), which are the most relevant for monitoring compliance in construction environments.

The slightly lower precision for the *background* class suggests that additional training data or refined annotations

could help reduce false detections in nonrelevant regions. Despite this limitation, the results confirm the model's suitability for real-time safety helmet detection and compliance monitoring at construction sites.

Fig. 9 illustrates how the F1-score varies with the confidence threshold for both the *protected* and *at risk* categories. At low confidence thresholds, the curves start at high values, reflecting strong initial detection capability.

The F1-score is defined as the harmonic mean of precision (P) and recall (R), as expressed in (1):

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (1)$$

where P denotes the ratio of correctly detected positive samples to the total predicted positives, and R represents the ratio of correctly detected positives to the total actual positives.

As the threshold increases, the F1-scores gradually decline, with a more pronounced drop observed for the *at risk* class once the value exceeds 0.6. Global performance, indicated by the blue curve, peaked at a F1-score of 0.89 when the confidence threshold was set to 0.435.

This value indicates the most favorable trade-off between precision and recall, ensuring reliable detection of individuals wearing helmets. These results highlight the importance of selecting an appropriate threshold to maximize the effectiveness of the system in real-world safety-monitoring applications.

As illustrated in Fig. 10, the precision–recall relationship is shown for the classes *protected* (helmet worn) and *at risk* (no helmet). The horizontal axis represents recall, while the vertical axis indicates precision, both ranging from 0 to 1.

The curve corresponding to the *protected* class reached a precision of 0.931, whereas the *at risk* class achieved 0.907. This result suggests that the model performs slightly better in identifying individuals wearing helmets than in detecting those at risk.

The overall mAP@0.5 score of 0.919 confirms a strong trade-off between precision and recall across classes, underscoring the robustness and general effectiveness of the YOLOv5s-based detection system for construction site monitoring.

Fig. 11 summarizes the evolution of key loss functions and evaluation metrics during the training and validation of the YOLOv5s model.

The *box_loss* for both training and validation decreased steadily across epochs, indicating a continuous improvement in bounding-box localization. Similarly, the *obj_loss* followed a downward trend, reflecting the model's increasing ability to identify the presence of objects. The *cls_loss* exhibited a sharp decline in the early epochs and stabilized near zero, which denotes enhanced accuracy in classifying detected objects.

B. F1-Score

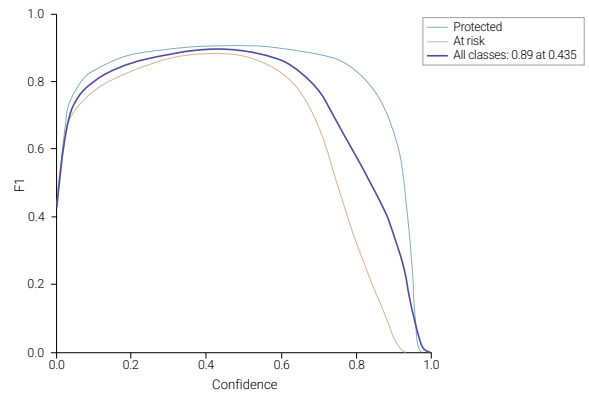


Fig. 9. F1-score.

C. Precision-Recall Curve

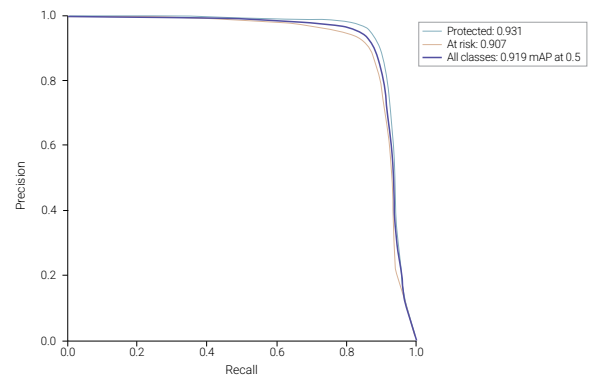


Fig. 10. Precision-recall curve.

D. Training and Validation Metrics

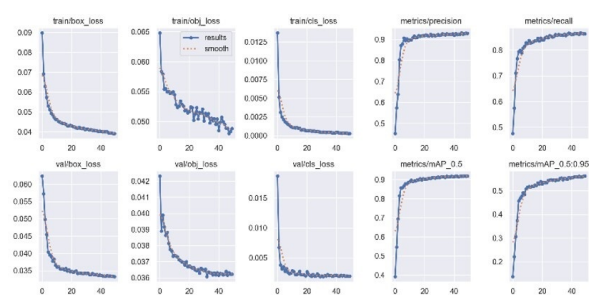


Fig. 11. Results of training and validation metrics.

Regarding performance indicators, both precision and recall improved progressively during training, reaching values close to 0.9. This trend confirms that the model became increasingly effective at producing correct detections while minimizing missed instances.

Furthermore, the mAP at an Intersection over Union (IoU) threshold of 0.5, as well as the mAP computed over the 0.5–0.95 range, increased consistently across training epochs.

These results demonstrate robust detection capability and strong generalization, underscoring the reliability of the trained YOLOv5s model in diverse operational scenarios.

E. System Demonstration

During the demonstration stage, the trained YOLOv5s model was deployed and evaluated under different real-world conditions. Validation included both video-based experiments—comprising one benchmark video and three recordings from active construction sites—and real-time testing with a webcam. These scenarios encompassed visitor entry, ongoing building activities, and worker assemblies, providing a representative sample of common construction site situations.

The deployment was carried out on a workstation equipped with an NVIDIA RTX 4060 GPU, achieving an average inference speed of 110 FPS. This ensured that the model remained resident in memory throughout the detection process, minimizing latency and enabling seamless real-time monitoring.

Figs. 12–15 illustrate representative frames captured during the demonstration phase, confirming the model’s capability to reliably distinguish between protected and at-risk individuals, labeled in Spanish as *protegido* (shown in green) and *riesgo* (shown in red) under operational conditions. These results validate the system’s suitability for live helmet compliance monitoring in construction environments.

For reproducibility and further development, the source code and implementation details of the proposed system are publicly available in a GitHub repository [10]. This allows interested researchers and developers to replicate the experiments, extend the work, or adapt the prototype to related applications.

IV. DISCUSSIONS

Compared to previous studies, this work demonstrates a tangible advancement in the detection of safety helmets within construction environments. Table III provides a comparative overview of key works in the field.

Hayat and Morgado-Dias [4] achieved 92.44% mAP with YOLOv5x but reported only 45 FPS, while our lighter YOLOv5s model reached a comparable 91.9% with more



Fig. 12. Real-world site visit scenario.

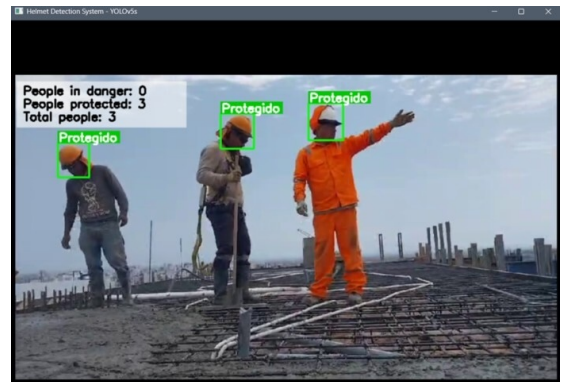


Fig. 13. Real-world construction scenario.

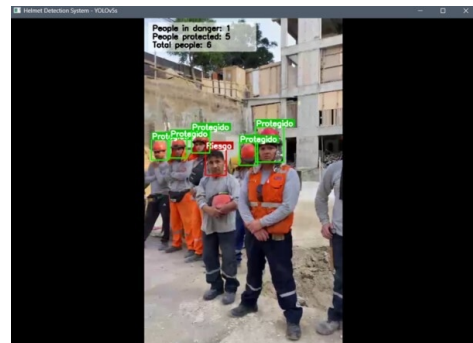


Fig. 14. Real-world worker assembly scenario.

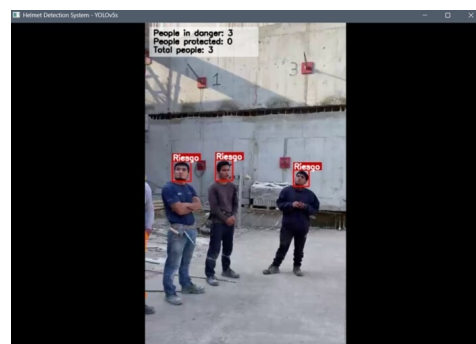


Fig. 15. Detection of at-risk workers.

than twice the processing speed (110 FPS). Qian and Wang [5] developed SHDet, optimized for embedded devices, reporting 92.2% mAP with extremely low inference time. An *et al.* [6] enhanced YOLOv5s with attention mechanisms, surpassing 93% mAP but at a higher computational cost. Li *et al.* [7] applied CNNs for helmet detection, showing good accuracy (~90%) but without validation under real-time conditions.

Unlike previous studies that focused primarily on accuracy improvements or embedded deployment, the novelty of this work lies in demonstrating that a standard YOLOv5s model, trained with a publicly available dataset, can achieve competitive accuracy while sustaining real-time inference at 110 FPS on accessible hardware. This highlights that reliable helmet compliance monitoring does not necessarily require complex model redesigns or high-end computing infrastructure, but can be deployed effectively using light-weight, scalable solutions.

Overall, our findings confirm that near state-of-the-art accuracy can be achieved with YOLOv5s while delivering superior inference speed on accessible hardware. Nevertheless, some limitations remain: false positives may occur when helmets are partially occluded or visually similar objects appear, and performance can degrade under extreme lighting or low-resolution inputs. Future work will evaluate newer architectures such as YOLOv8 or RT-DETR to determine whether accuracy gains can be achieved without sacrificing real-time performance.

In summary, this study provides a robust and efficient solution for real-time helmet compliance monitoring, striking a practical balance between detection accuracy, speed, and deployment feasibility.

V. CONCLUSION

This work presented a computer vision system for helmet compliance detection on construction sites using the YOLOv5s algorithm. The results demonstrated that the model provides accurate, real-time monitoring while maintaining high inference speed, making it a practical solution for improving occupational safety and reducing accident risks. By automating the supervision process, the system enhances regulatory compliance and alleviates the need for continuous manual oversight, contributing to more efficient site management.

The main contribution of this study is the demonstration that YOLOv5s can deliver both high detection accuracy and real-time performance on widely available hardware, offering a cost-effective and scalable alternative for construction site safety monitoring.

Future research will focus on extending detection capabilities to additional personal protective equipment such as vests, goggles, and gloves, thus broadening its safety coverage. Furthermore, the system will be tested under

TABLE III
COMPARISON OF RELATED STUDIES ON SAFETY
HELMET DETECTION

Study	Model	mAP@0.5	FPS
Hayat and Morgado-Dias [4]	YOLOv5x	92.44%	45
Qian and Wang [5]	SHDet	92.2%	-
An <i>et al.</i> [6]	YOLOv5s+	93.5%	70
Li <i>et al.</i> [7]	CNN-based	~90%	-
This work (2025)	YOLOv5s	91.9%	110

challenging environmental conditions, including low-light and adverse weather scenarios, to validate robustness. Integration with CCTV infrastructures and IoT-based alert mechanisms is also envisioned to support large-scale, real-world deployments.

REFERENCES

- [1] U.S. Bureau of Labor Statistics, "Census of fatal occupational injuries summary, 2023," U.S. Department of Labor, Washington, DC, USA, Rep., Dec. 2024. [Online]. Available: <https://www.bls.gov/news.release/cfoi.nr0.htm>
- [2] Occupational Safety and Health Administration, "OSHA announces switch from traditional hard hats to safety helmets to protect agency employees from head injuries better," U.S. Department of Labor News Release, Dec. 11, 2023. [Online]. Available: <https://www.osha.gov/news/newsreleases/trade/12112023>
- [3] M. Bottlang, G. DiGiacomo, S. Tsai, and S. Madey, "Effect of helmet design on impact performance of industrial safety helmets," *Heliyon*, vol. 8, no. 8, Art. no. e09962, Aug. 2022, doi: <https://doi.org/10.1016/j.heliyon.2022.e09962>
- [4] A. Hayat and F. Morgado-Dias, "Deep learning-based automatic safety helmet detection system for construction safety," *Appl. Sci.*, vol. 12, no. 16, Art. no. 8268, Aug. 2022, doi: <https://doi.org/10.3390/app12168268>
- [5] Y. Qian and B. Wang, "A new method for safety helmet detection based on convolutional neural network," *PLoS One*, vol. 18, no. 10, Art. no. e0292970, Oct. 2023, doi: <https://doi.org/10.1371/journal.pone.0292970>
- [6] Q. An, Y. Xu, J. Yu, M. Tang, T. Liu, and F. Xu, "Research on safety helmet detection algorithm based on improved YOLOv5s," *Sensors*, vol. 23, no. 13, p. 5824, Jun. 2023, doi: <https://doi.org/10.3390/s23135824>
- [7] Y. Li, H. Wei, Z. Han, J. Huang, and W. Wang, "Deep learning-based safety helmet detection in engineering

- management based on convolutional neural networks,” *Adv. Civ. Eng.*, vol. 2020, no. 6, pp. 1–10, Sep. 2020, doi: <https://doi.org/10.1155/2020/9703560>
- [8] Y. Li, C. Ma, L. Li, R. Wang, Z. Liu, and Z. Sun, “Lightweight tunnel obstacle detection based on improved YOLOv5,” *Sensors*, vol. 24, no. 2, p. 395, Jan. 2024, doi: <https://doi.org/10.3390/s24020395>
- [9] HelmetWearingData, “YOLOv5 hard hat detection dataset v2,” Roboflow Universe, 2023. [Online]. Available: https://universe.roboflow.com/helmetwearingdata/yolov5-hard_hat_detection/dataset/2
- [10] P. Suarez and M. Paz, “Helmet Detection Project,” GitHub Repository, 2025. [Online]. Available: https://github.com/pierosuarez29/Helmet_Detection_Project