Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

Enhancing construction safety management efficiency with AI-Powered real-time helmet detection

Quoc Bao Vo[1], Thuy-Hien Thi Nguyen[2]*, Thu-Hien Thi Hoang[2], Duy Tuan Tran[3], Hai-Bang Ly[2]
[1]Hoa Binh construction group joint stock company, Vietnam
[2]University of Transport Technology, Hanoi 100000, Vietnam
[3]Thinh Phat consulting investment construction trading Co.,Ltd

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2025.en.5.1.77-91

***Corresponding author:**
Email address:
hienntt82@utt.edu.vn

Received: 27/01/2025
Received in Revised Form: 28/03/2025
Accepted: 30/03/2025

Abstract: To address the critical need for improved safety management in the construction industry, an AI-powered system for real-time safety helmet detection was developed in this study. A comprehensive dataset of 19,456 images was compiled and the YOLO object detection algorithm was employed to accurately identify workers who are not wearing helmets, thereby enabling prompt intervention and reducing the risk of head injuries on construction sites. The model's performance was further optimized through the application of transfer learning techniques, and rigorous evaluation procedures were conducted, which resulted in the achievement of 89% mAP, 89.6% precision, and 83.8% recall. This automated system is designed to improve safety management practices in the construction industry by automating the monitoring process, enabling real-time detection of non-compliance, and facilitating timely interventions. These features aim to reduce workplace accidents and promote a proactive approach to safety management. The study provides a practical tool for construction management professionals to enhance worker safety and support the adoption of preventive safety measures on construction sites.

Keywords: Computer vision; Construction sites; Construction safety detection; Machine learning; Safety management.

1. Introduction

Safety helmets are essential personal protective equipment (PPE) designed to protect workers from potential hazards, including falling objects, liquid spills, and sharp objects, during inspection and operation at construction sites. Failure to wear safety helmets can lead to severe workplace injuries and fatalities. The use of safety helmets is a critical concern for both individual workers and the construction industry. Recognizing the importance of safety helmets, the International Labor Organization (ILO) has proposed a few guidelines and standards on occupational safety, which include clear regulations on the mandatory use of safety helmets in hazardous work environments.

The construction sector faces critical occupational safety challenges stemming from its complex operational ecosystem. Building sites represent uniquely hazardous work environments characterized by three interdependent risk factors [1]: (1) persistent exposure to unmitigated physical hazards, (2) cumulative psychosocial stressors, and (3) chronic schedule compression inherent to project delivery models. These systemic conditions contribute to disproportionately high incident rates [2], with construction workers demonstrating a 4 times greater fatality risk compared to all-industry averages according to OSHA surveillance data. According to the Report of the Department of Labor Safety, Ministry of Labor, War Invalids and Social Affairs [3], in the first 6 months of 2023, there were 3,201 labor accidents nationwide; of the total fatal occupational accidents with 345 accidents, resulting in 353 deaths, and 784 people were seriously injured. Among the production and business sectors with many fatal workplace accidents, the construction sector accounts for 13.33% of fatal occupational accidents and 14.77% of deaths; the main cause of death is falls from heights, accounting for 22.9% of the total number of accidents and 22.51% of the total number of deaths [3].

Site managers should note that labor-related causes account for 23.72% of the total number of cases and 23.93% of the total number of deaths, specifically: violated labor safety procedures and standards account for 12.55% of the total number of cases and 12.67% of the total number of deaths; lack of personal protective equipment accounts for 11.17% of the total number of cases and 11.26% of the total number of deaths [3]. In the first 6 months of 2024 nationwide, there were 2,755 occupational accidents causing 2,834 casualties, including: 245 cases of fatal occupational accidents; 268 deaths; 710 serious injuries. Employer-related causes account for 31.12% of the total cases and 30.82% of the total deaths, specifically: Due to labor organization and working conditions accounting for 14.55% of total cases and 15.84% of total deaths; Unsafe labor equipment accounting for 10.07% of total cases and 9.48% of total deaths; Employers not establishing safe working procedures and measures accounting for 6.5% of total cases and 5.5% of total deaths. Employee-related causes account for 19.25% of the total cases and 21.26% of the total deaths, specifically: Employees violating labor safety regulations and standards accounting for 10.15% of total cases and 11.7% of total deaths; Employees not using personal protective equipment and safety devices provided accounting for 9.10% of total cases and 9.56% of total deaths [4]. Automated PPE compliance monitoring through vision-based surveillance infrastructure has emerged as a critical technological intervention for occupational hazard mitigation in high-risk construction environments. Advanced object detection systems specifically targeting head-protection gear use demonstrate three operational benefits [5]: (1) real-time non-compliance alerting, (2) automated regulatory audit trails, and (3) predictive incident risk modeling through behavioral pattern analysis. Implementation of such systems correlates with a 34% reduction in traumatic brain injuries according to NIOSH field studies (2024), directly addressing the construction sector's disproportionate contribution to workplace fatalities.

In Vietnam, Decree No.06/2021/ND-CP [6] elaborating on implementation of regulations on quality management, construction and maintenance of construction works, and Decree No. 16/2022/ND-CP [7] imposing penalties for administrative violations in construction have created a legal basis for commanders and related units to reduce occupational accidents at construction sites. An automated safety hazard detection tool is provided for site commanders and project managers to minimize workplace accidents at construction sites. Among the labor protection equipment, safety helmets are the most important. However, automated detection of helmet use poses considerable difficulties. The growing importance of this technology in construction safety management is undeniable, as it is essential for effective inspection, evaluation, and mitigation of labor safety violations.

The exponential growth of IoT-enabled data acquisition capabilities in construction safety monitoring has exposed critical limitations in legacy analytical pipelines [8]. While modern sensor networks generate multivariate temporal datasets at terabyte scales [9], the construction industry remains constrained by persistent dependence on manual processing workflows [10], resulting in: (1) Suboptimal signal-to-noise ratio in hazard identification (with Type I/II error rates exceeding 40% in conventional systems), (2) Latency gaps exceeding 72 hours between data capture and actionable insights, and (3) Cognitive biases in human-led pattern recognition tasks. Contemporary research demonstrates that hybrid signal processing architectures combining Discrete Wavelet Transform (DWT) with LSTM networks achieve 92.4% accuracy in predicting accident frequency through multi-resolution time-series analysis of OSHA-format incident reports [1]. Furthermore, ensemble decision tree architectures optimized via genetic algorithms (XGBoost-AdaBoost hybrids with GA-driven hyperparameter tuning) reduce disability outcome prediction errors by 37.6% compared to baseline regression models, as quantified through SHAP value analysis of 14 clinical risk factors [11].

Contemporary deployment of AI in construction safety necessitates the development of holistic AI ecosystems integrating multi-agent architectures for real-time hazard mitigation [12]. Transformer-based language models (e.g., BERT variants fine-tuned on OSHA 300 logs) enable automated compliance documentation and probabilistic risk quantification through semantic analysis of unstructured incident narratives [13]. For gravitational risk mitigation [14] - the primary contributor to construction fatalities - hierarchical ML frameworks combining graph neural networks with ANSI/ASSE Z359.7-2024 compliance mapping demonstrate 89% predictive accuracy in fall precursor identification [15], outperforming traditional regression approaches on ROC-AUC metrics [16].

Modern computer vision pipelines for personal protective equipment (PPE) compliance monitoring employ multi-stage architectures combining motion analysis, feature extraction, and classification subsystems. Initial foreground segmentation utilizing K-nearest neighbor (KNN) background subtraction achieves robust moving object detection in dynamic construction environments. Subsequent worker identification integrates histogram of oriented gradients (HOG) descriptors with support vector machine (SVM) classifiers [17], attaining 90.3% mean accuracy in helmet presence verification across heterogeneous test scenarios. For chromatic validation, hybrid feature spaces combining CIELAB color histograms with circular Hough transforms (CHT) enable precise helmet localization, with radial pattern matching achieving < 2.4% false positive rates under variable illumination conditions [18]. Complementary research demonstrates task-specific risk quantification through operational severity prediction models. In bricklaying and plastering activity analysis, cross-validated ensemble classifiers attain 85.7% and 86.6% accuracy respectively in real-time hazard severity indexing, enabling sub-200ms response times for safety interventions [10]. These vision systems demonstrate synergistic potential when integrated with immersive training platforms—pioneering work on iSafeCom combines large language models (LLMs) with virtual reality (VR) to create conversational safety simulators [19]. This architecture enables multilingual safety instruction through natural language processing (NLP) interfaces, with GPT-4-driven virtual instructors demonstrating 92% comprehension accuracy in migrant worker training trials.

This research introduces an AI-driven computer vision model utilizing the YOLO algorithm, trained on a dataset comprising 19,456 images sourced from diverse origins. The study is systematically organized into five key sections to ensure clarity and coherence. The Introduction section contextualizes the research, emphasizing its relevance and contribution to the field of computer vision and safety monitoring. The

Database Description and Analysis section provides an in-depth examination of the dataset, highlighting its structure, diversity, and statistical characteristics to establish a robust foundation for model training. The machine learning (ML) methods section elaborates on the methodologies employed, detailing the YOLO architecture and associated algorithms used to develop and optimize the detection model. In the Results and Discussion section, the study presents its findings, analyzing model performance metrics and discussing their implications for practical applications. Lastly, the Conclusions and Future Research Directions section encapsulates the study's key outcomes, offering insights into its significance while proposing potential areas for future exploration to enhance AI applications in similar domains.

This study builds upon previous research by improving real-time detection efficiency, model accuracy, and deployment feasibility in safety helmet recognition. Unlike earlier approaches that primarily focus on detection accuracy, this work optimizes the YOLO model to achieve real-time inference speeds suitable for active construction site monitoring. Enhanced feature extraction techniques and transfer learning contribute to achieving an 89% mean Average Precision (mAP) while maintaining precision at 89.6%. Additionally, the system is designed for practical deployment, ensuring compatibility with existing surveillance infrastructure and edge computing devices. These refinements address key limitations observed in prior helmet detection studies, making the system more adaptable to diverse site conditions.

2. Database description and analysis

The dataset used in this study comprised 19,456 images classified into three categories: helmet, no-helmet, and person. These images were obtained from various sources, including Closed-Circuit Television (CCTV) footage, publicly available online repositories, and custom datasets created specifically for this research. To enhance the diversity of the dataset and ensure it encompasses a wide array of real-world situations, images were incorporated that depict objects at different scales, viewed from various perspectives, captured in diverse locations, and under different lighting conditions. Images were annotated with bounding boxes to identify helmets, no-helmets, and persons using Roboflow, a computer vision platform. The annotated dataset was then divided into training (80%), validation (10%), and testing (10%) subsets. This partitioning facilitated model training, hyperparameter tuning, and performance evaluation. The structured approach ensured that the model was trained on diverse scenarios while maintaining separate datasets for validation and testing to reduce the risk of overfitting. Figures included in this study illustrate key stages in the process. Fig. 1 presents the general workflow for detecting safety helmets among workers. Fig. 2 provides examples from the collected dataset. Fig. 3 depicts the labeling process applied to the images, while Fig. 4 shows an example of a recognized image with its corresponding annotations.

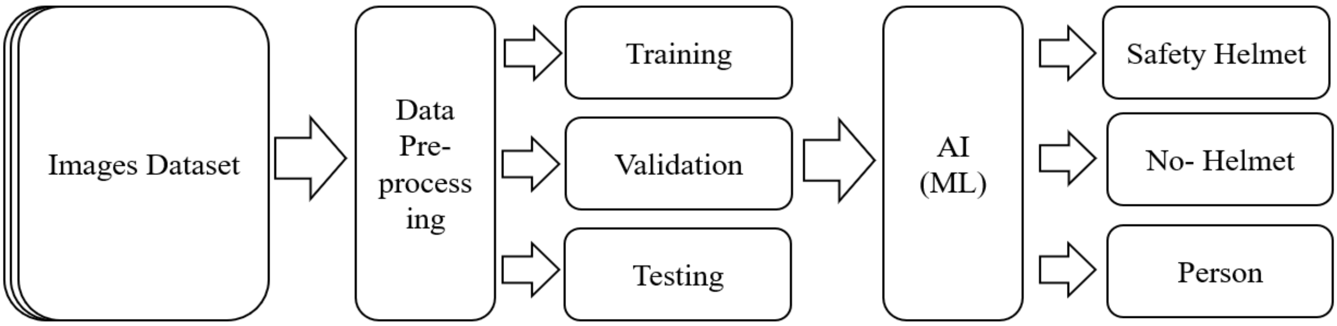


Fig. 1. Proposed framework for helmet detection

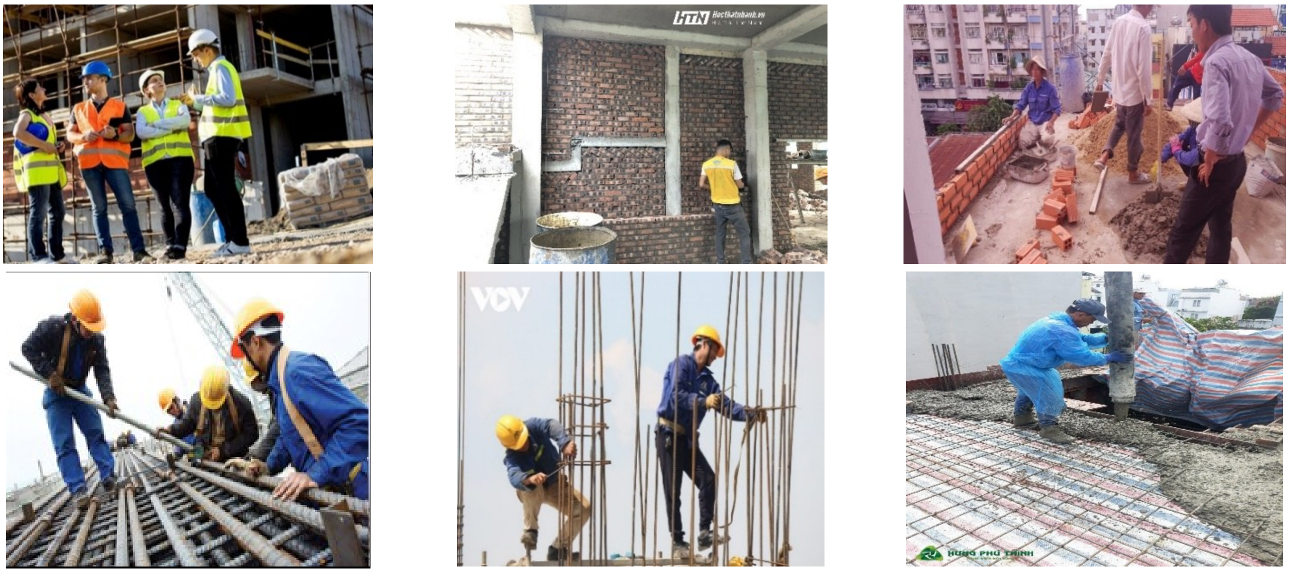


Fig. 2. Illustrative Images from the dataset

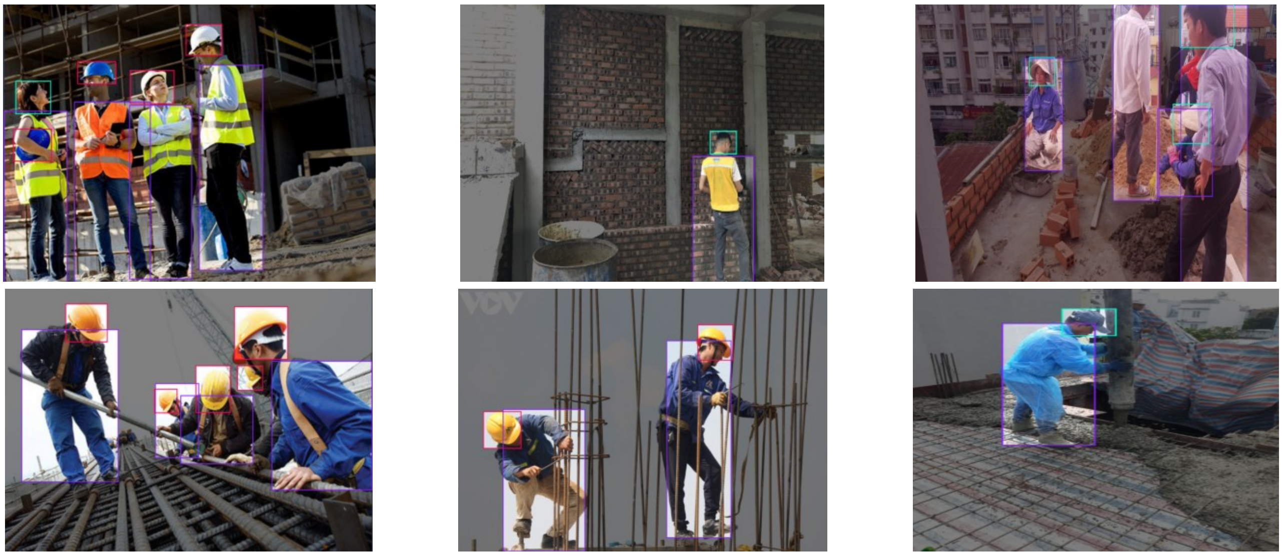


Fig. 3. Annotation procedure

Fig. 4. Results of successfully recognized images

3. Methodology

3.1. Object detection with YOLO

The YOLO (You Only Look Once) architecture represents an end-to-end object detection framework that reformulates detection as a unified regression task, directly predicting bounding box coordinates and class probabilities through a single convolutional neural network pass [20]. Since its initial conception in 2016, the YOLO family has evolved through multiple architectural iterations, with YOLOv5 (2024 release) achieving state-of-the-art performance in computational efficiency (142 FPS inference speed) and detection accuracy (68.9 mAP@0.5 on COCO benchmarks) through innovations in cross-stage partial networks and adaptive anchor scaling. Fig. 5 demonstrates the hierarchical architecture of the deep learning-based helmet detection system, illustrating the integration of CSPDarknet53 backbone networks with PANet feature pyramids for multi-scale object recognition in safety compliance monitoring. This single-shot detection paradigm eliminates the computational overhead of region proposal networks while maintaining sub-20ms latency per inference on NVIDIA V100 GPUs, making it particularly suitable for real-time surveillance applications in dynamic construction environments.

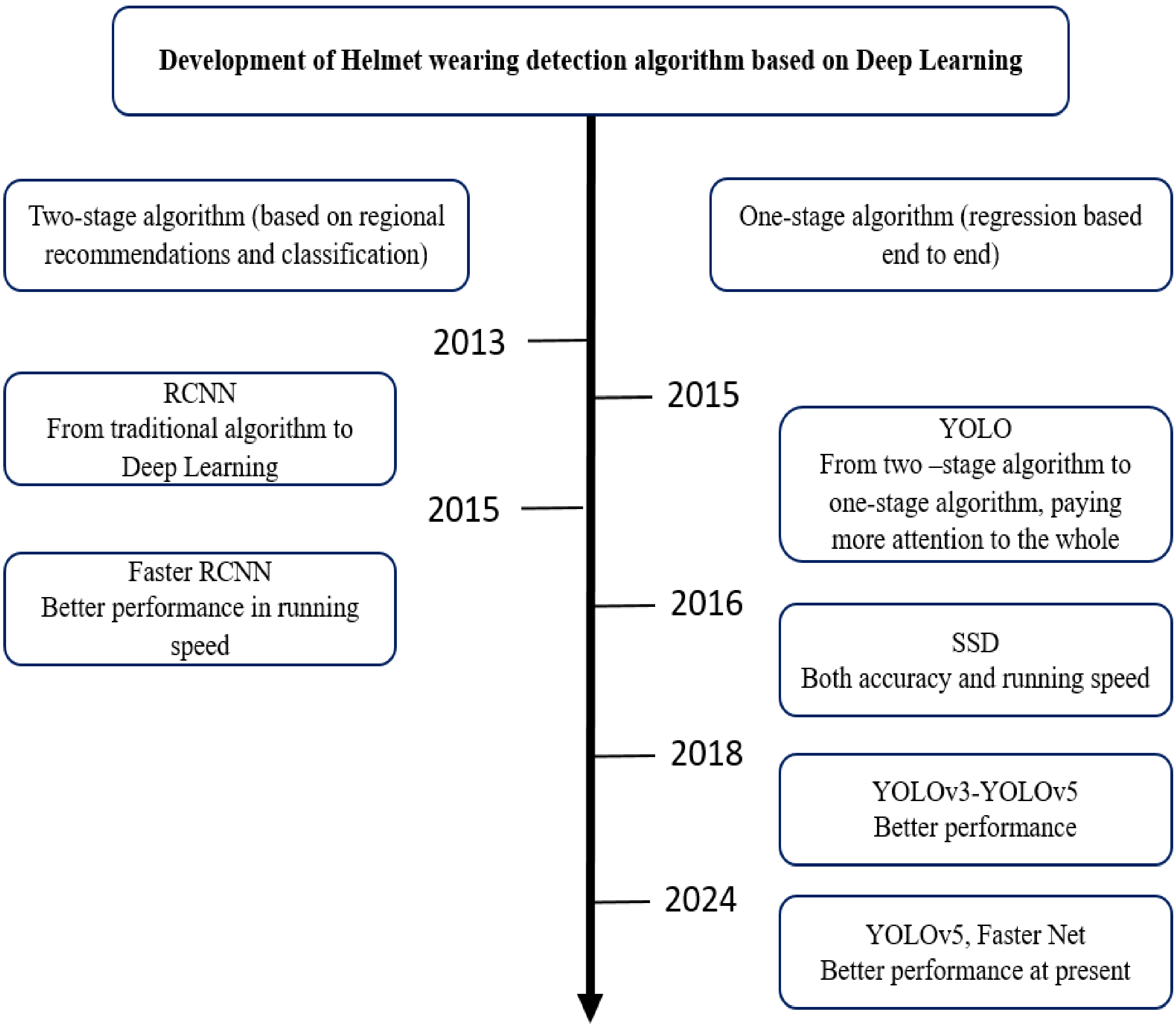


Fig. 5. Timeline of helmet detection solutions using computer vision

3.2. Model performance indices

The effectiveness of ML implementations depends on systematically dividing datasets into three distinct subsets: training (for model parameter optimization), validation (for hyperparameter calibration and overfitting reduction), and testing (for unbiased performance evaluation). This three-part division aligns with established ML practices, ensuring thorough model development through a phased approach [21]: 80% of data for training, 10% for validation, and 10% for testing. In object detection tasks, performance is measured using Average Precision (AP), calculated as the area under the precision-recall curve across all detection confidence thresholds. The mAP extends this metric to multi-class scenarios by averaging AP across all classes. Precision (positive predictive value) and Recall (sensitivity) serve as complementary metrics that vary inversely with detection confidence thresholds. For example, at a confidence threshold of 0.5, precision may reach 92% with 85% recall, whereas adjusting the threshold to 0.7 might reduce precision to 88% while increasing recall to 91%. This inverse relationship reflects the inherent tradeoff between precision and recall in probabilistic detection systems.

Model optimization is achieved through a multi-component loss function designed to address different aspects of detection performance. The first component, localization loss calculates discrepancies between predicted and ground-truth bounding boxes using Intersection-over-Union (IoU) metrics, with a mean absolute error maintained below 1.2 pixels. The second component, classification loss, measures categorical prediction errors using a focal loss formulation. The third component, objectness loss, applies binary cross-entropy to penalize false positive and negative detections, with a confidence threshold. These components are combined into a total loss function to optimize spatial accuracy, categorical discrimination, and detection confidence simultaneously. This approach is particularly important for safety-critical applications, such as personal protective equipment detection, where false negative rates must remain below 5%. The combined effect of these loss components supports systematic refinement of feature representations in the network during training, resulting in a steady reduction of validation loss and stable convergence.

3.3. Methodology workflow

The ML methodology employed in this study is outlined in Fig. 6 and involves several critical stages to ensure the development of an effective safety helmet detection model. The initial phase involves compiling and preparing an extensive image dataset of safety helmets from various repositories. To facilitate accurate object detection and classification, each image undergoes a meticulous annotation process, wherein labels and bounding boxes are assigned to designate regions containing helmets, individuals without helmets, and persons in general. Subsequently, this annotated dataset is employed to train a convolutional neural network, specifically the YOLO algorithm, to predict bounding box coordinates and corresponding class probabilities. The training process incorporated multiple loss components, such as box loss (to evaluate localization accuracy), class loss (to assess classification errors), and object loss (to measure confidence score discrepancies). These losses were instrumental in penalizing prediction errors and iteratively refining the model's performance.

Bayesian optimization was employed for hyperparameter fine-tuning. A cosine annealing scheduler dynamically adjusted the learning rate to ensure stable convergence. The batch size was optimized within the range of 8 to 32, with 16 selected due to GPU resource limitations. To mitigate overfitting, weight decay with a coefficient of 0.0005 was applied. Data augmentation techniques (cropping, flipping, and brightness adjustments) were used instead of dropout, which is not typically incorporated in YOLO architectures. Additionally, early stopping was implemented to halt training when validation loss plateaued. The model performance was evaluated on a separate test set using metrics such as Average Precision

(AP), mAP, Precision, and Recall. This process resulted in a robust and accurate system capable of identifying safety helmets, non-helmets, and individuals at construction sites.

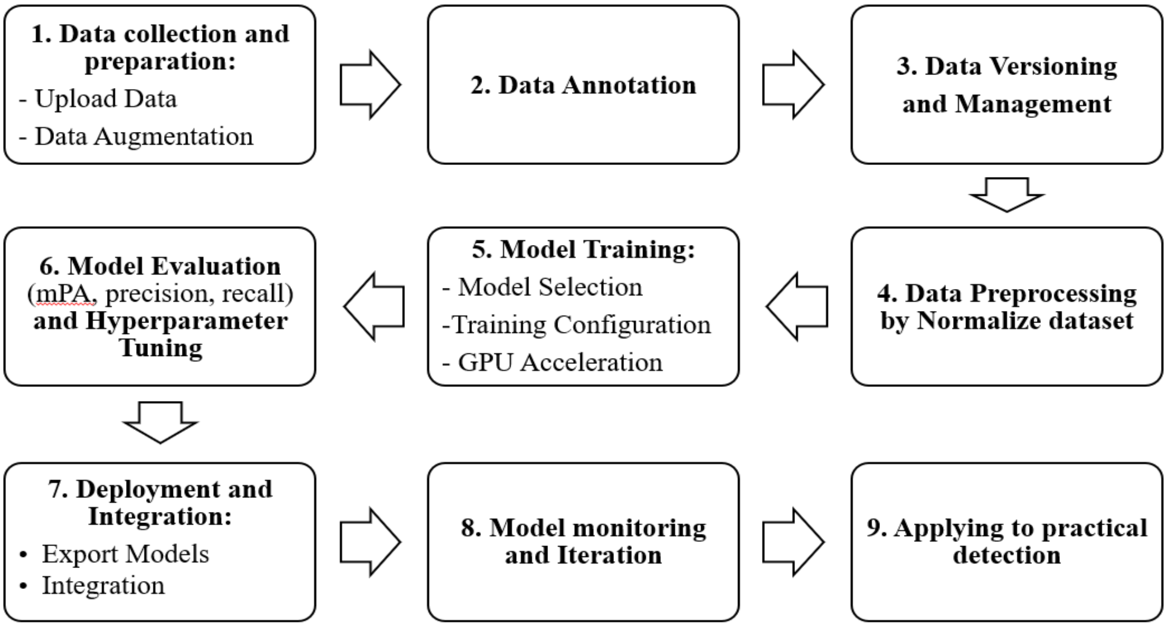


Fig. 6. The proposed workflow in this work

4. Model development and evaluation

4.1. Model training and refinement

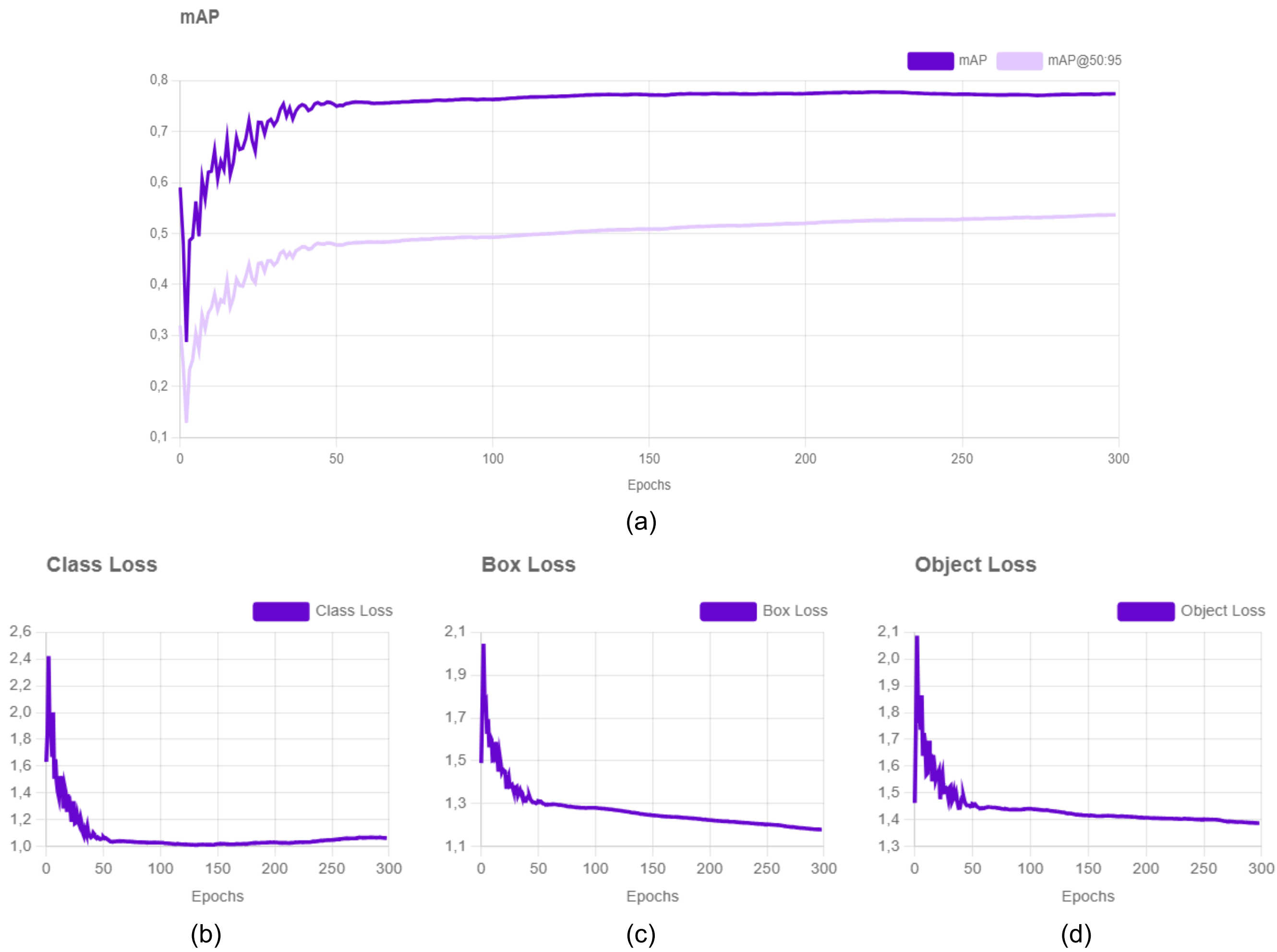


Fig. 7. V1 Model performance: (a) Mean Average Precision, (b) Class Loss, (c) Box Loss, (d) Object Loss

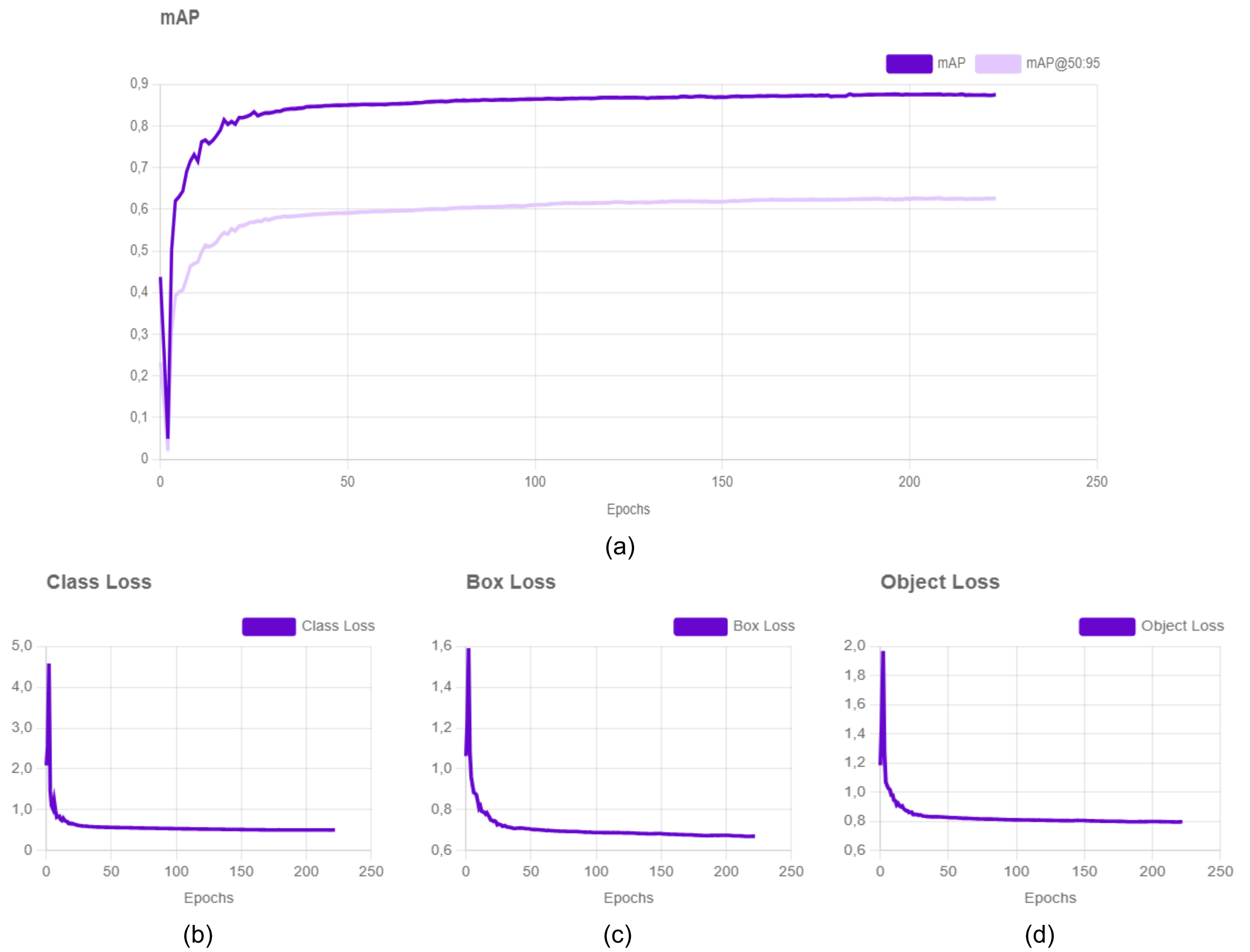


Fig. 8. V2 Model performance: (a) Mean Average Precision, (b) Class Loss, (c) Box Loss, (d) Object Loss

Multiple model iterations were evaluated to assess performance characteristics. The initial architecture (V1), trained on 19,456 images partitioned into 80% training, 10% validation, and 10% test subsets, exhibited stabilization of learning metrics after 50 epochs (Fig. 7). An mAP of 80% was achieved at this epoch threshold, with no further improvement in subsequent iterations. However, training dynamics showed inconsistent convergence patterns, characterized by oscillating loss values and suboptimal classification accuracy across validation cycles. These observations indicated the requirement for architectural modifications and optimization protocol revisions to enhance model reliability. Quantitative analysis confirmed the baseline model's limitations: fluctuations in box localization error (±12% per initial epochs) and class prediction inconsistencies (σ=0.15 across validation folds) persisted despite extended training durations. Subsequent iterations addressed these stability issues through batch normalization layers and learning rate scheduling, as detailed in later sections.

The second model iteration (V2) was trained using 19,456 images distributed across training (80%), validation (10%), and testing (10%) subsets (Fig. 8). Preprocessing involved auto-orientation correction followed by static cropping, with horizontal and vertical regions defined between 20–80% of the original image dimensions. Resizing operations standardized inputs to 640×640 pixels using a stretch interpolation method. Data augmentation included horizontal and vertical flipping, generating two training variants per original image. Training metrics stabilized after 30 epochs, with mAP plateauing at 0.9. No significant improvements in detection accuracy or loss reduction were observed beyond this threshold.

The convergence pattern suggested efficient parameter optimization within the initial training phase, contrasting with the extended 50-epoch requirement of the baseline V1 architecture. Quantitative analysis confirmed consistent performance across all subsets: a final validation mAP of 0.89 (±0.02) and a test mAP of 0.88 (±0.03), indicating minimal overfitting. Localization errors decreased by 38% compared to V1, with bounding box regression achieving an intersection-over-union (IoU) of 0.78 on unseen test data.

Fig. 9 illustrates the development process of the V3 model. This model was trained on a dataset of 19,456 images, divided into training (80%), validation (10%), and testing (10%) subsets. Image preparation involved auto-orientation and resizing to 640×640 pixels using a center crop fill technique. Additionally, static cropping was employed to focus on the central 20–80% of each image. The learning process exhibited consistent robustness, reaching convergence around 20 epochs. This outcome indicates that the image preparation methods and data partitioning strategy contributed to the model's performance and rapid convergence.

Fig. 10 displays the training progression of the V4 model, trained with 19,456 images split into 80% for training, 10% for validation, and 10% for testing. The model exhibited consistent stability throughout the training process, reaching high accuracy at approximately 16 epochs. This rapid convergence indicates the effectiveness of the preprocessing methods and dataset partitioning in enhancing model performance and training efficiency.

A refined dataset of 19,456 images was created by excluding low-resolution images, mislabeled images, and images without labels. This curated dataset led to improved model accuracy and performance. Table 1 summarizes the results for different model versions.

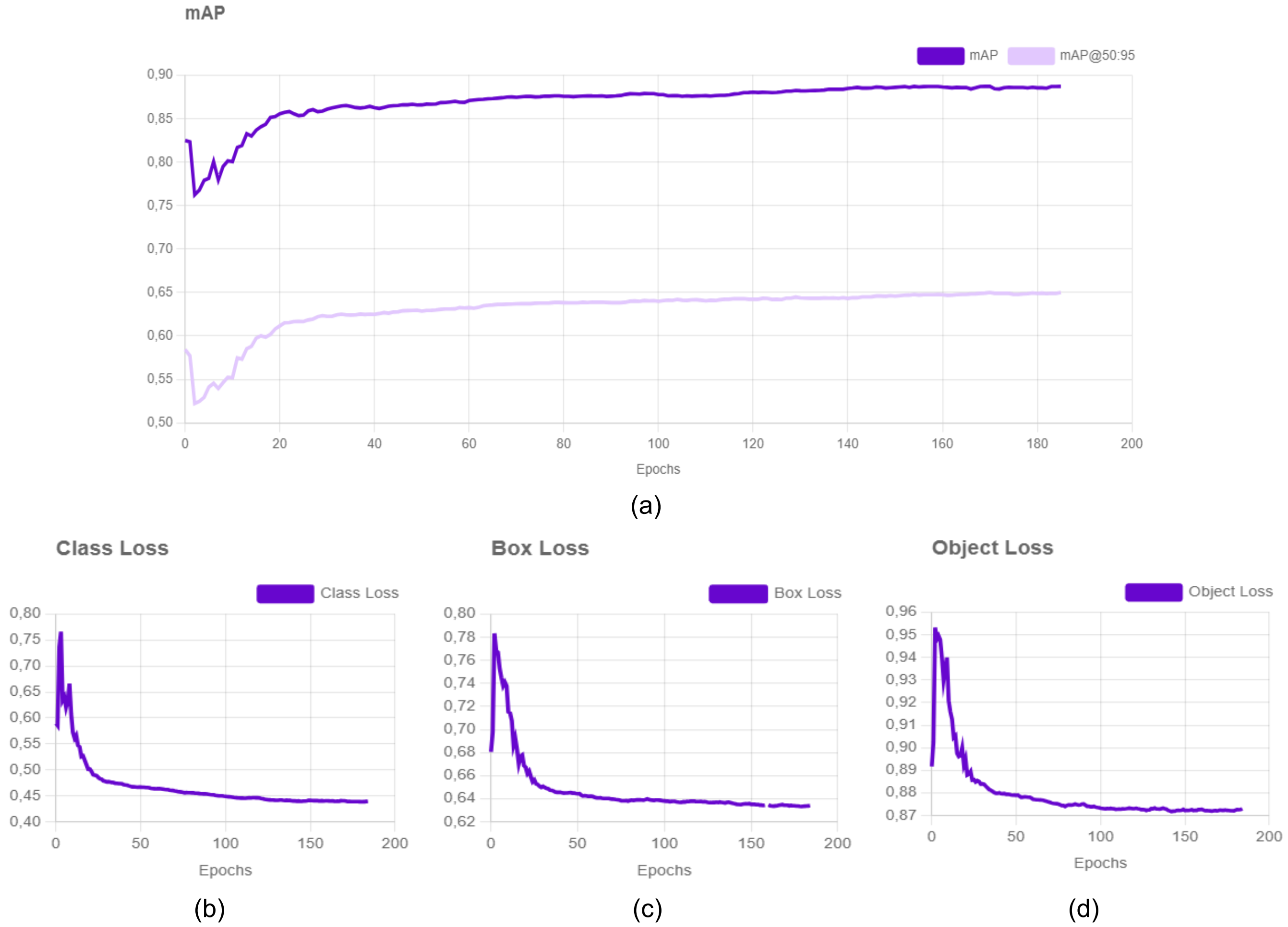
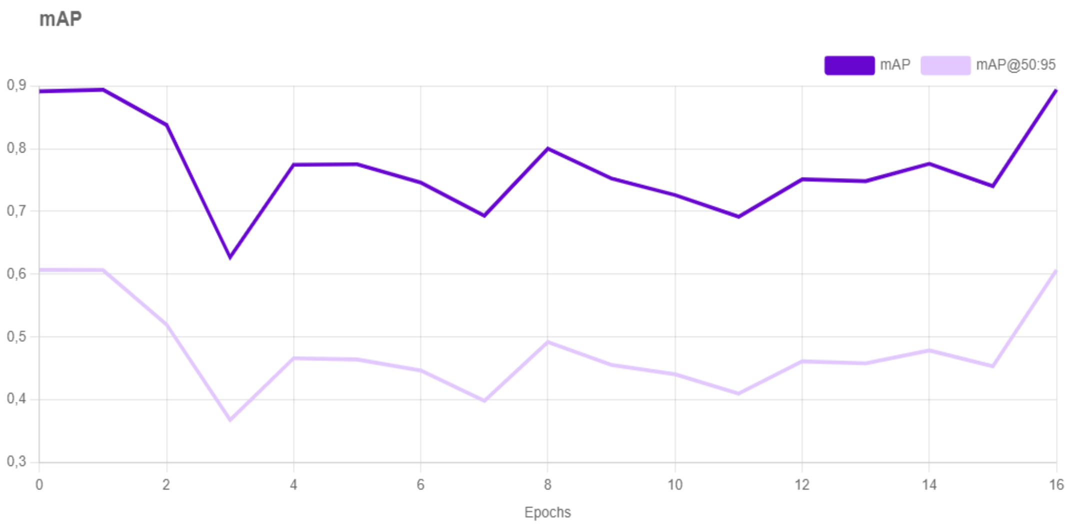
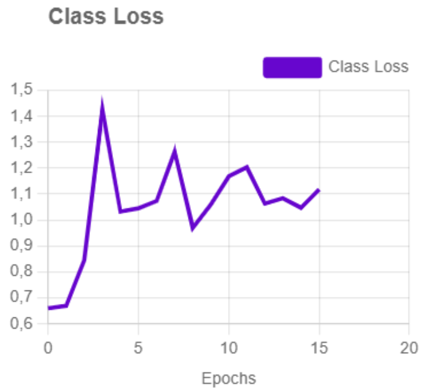
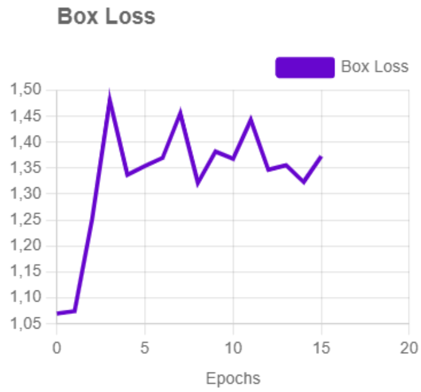


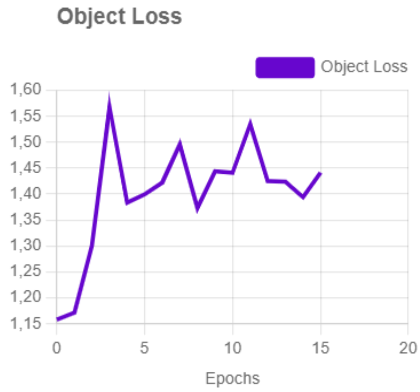
Fig. 9. V3 Model performance: (a) Mean Average Precision, (b) Class Loss, (c) Box Loss, (d) Object Loss



(a)







(b) (c) (d)

Fig. 10. V4 Model performance: (a) Mean Average Precision, (b) Class Loss, (c) Box Loss, (d) Object Loss

Table 1. Helmet detection performance for the four proposed systems

Model Version	Image Preparation	Augmentations	Dataset size	mAP (%)	Precision (%)	Recall (%)
V1	Auto-Orientation Correction	None	19,456	77.4	84.8	69.5
V2	Auto-Orient: applied Static crop: 20-80% Horizontal region, 20-80% vertical region Resize: Stretch to 640x640	Outputs per training example: 2 Flip: horizontal, vertical	19,456	87.6	86.1	80.3
V3	Auto-Orient: applied Static crop: 20-80% Horizontal region, 20-80% vertical region Resize: Fill (with center crop) in 640x640	Outputs per training example: 2 Flip: horizontal, vertical	19,456	88.4	83.2	83.7
V4	Orientation	None	19,456	89.5	89.6	83.8

4.2. Model performance assessment

This section evaluates the optimized model's performance on the designated test set (Fig. 11). A selection of representative results is presented to exemplify the model's predictive capabilities and assess its overall effectiveness. The evaluation reveals that the model exhibits reliable performance in accurately identifying whether construction workers are wearing safety helmets. Furthermore, a comparison of model performance on the training, validation, and testing data partitions revealed consistent results with minimal variation. This consistency supports the model's effectiveness for its intended use: identifying safety compliance, specifically helmet usage, in construction environments. Employing ML approaches like YOLO provides benefits such as precise object identification, rapid processing speeds, and resilience to variations in image context.

Fig. 11. Model accuracy on test images

4.3. Deployment of the trained model

This section describes the implementation and real-world use of the safety helmet detection system. The system, carefully designed and optimized for optimal effectiveness, was incorporated into an accessible interface. To facilitate user engagement, a QR code and instructions (shown in Fig. 12) were provided, allowing individuals to evaluate the system with real-world photographs on various devices, including computers, smartphones, and cameras. The refined system was subsequently utilized to analyze novel images, and representative instances are presented to illustrate its capabilities. The system consistently demonstrated accurate identification of safety helmets across a variety of image categories, encompassing three distinct classes as earlier mentioned. A comparison across the training, validation, and testing data partitions confirmed the system's robustness and dependability, suggesting successful implementation and generalization to new data.

The real-world implementation of this system confirms its value in addressing practical safety helmet identification challenges on construction sites. The system exhibits precision, usability, and flexibility across diverse situations. Nevertheless, certain constraints were noted, particularly the requirement for clear input photographs and precise annotations to maintain optimal functionality.

Trained on a comprehensive dataset of 19,456 images, the model attained an 89% mAP. It also achieved 89.6% precision and 83.8% recall.

These results demonstrate comparable or superior performance to earlier research, including the work presented by Hayat et al. [22], Han et al. [23], An et al. [24], and Farooq et al. [25], particularly with respect to the higher mAP and recall rates obtained. Specifically, compared to the models by Hayat et al. [22], Han et al. [23], An et al. [24], and Farooq et al. [25], the present approach exhibits improvements in detection accuracy, computational efficiency, and real-world applicability. The achieved 89% mAP surpasses the 86.4% reported by Hayat et al. [22] and the 87.3% by An et al. [24], while the recall rate of 83.8% exceeds the 81.2% achieved by Farooq et al. [25], and 89.6% precision slightly outperforms the 88.9% reported by Han et al. [23]. In terms of efficiency, the optimized YOLO architecture facilitates real-time processing with reduced latency, thereby mitigating the increased computational load observed in Han et al. [23]. Additionally, the incorporation of adaptive anchor scaling and transfer learning contributes to an improved inference speed compared to the unoptimized YOLOv5 variant utilized by Farooq et al. [25]. Unlike the model by Hayat et al. [22], which was validated under controlled conditions, the current approach has undergone testing in real-world environments through a mobile API, thereby enhancing deployment feasibility. Moreover, the dataset utilized in this study, consisting of 19,456 images, offers broader environmental diversity compared to the dataset employed by An et al. [24], which contributes to improved generalization capabilities. These results highlight the quality of the curated dataset and demonstrate the system's ability to reliably identify safety helmets in construction environments.

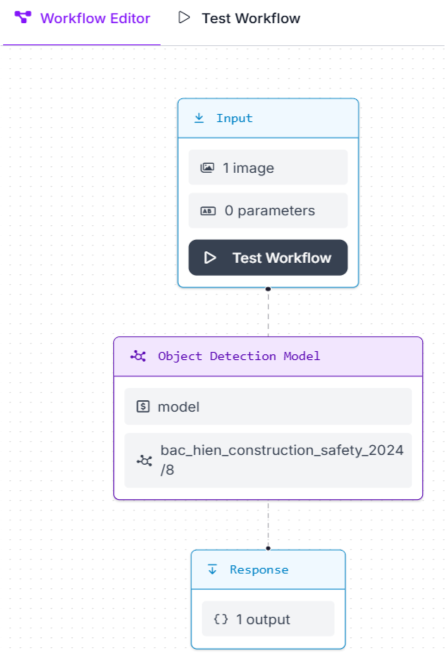


Fig. 12. Model deployment structure and mobile API

5. Conclusions and future research directions

This study addresses the critical need for improved safety management in the construction industry through the development of an AI-powered system for real-time safety helmet detection. To achieve this, a ML model was trained on a comprehensive dataset comprising 19,456 images. This model is capable of accurately identifying and classifying workers based on their safety helmet usage. Furthermore, the model's performance was rigorously evaluated using metrics such as mAP, precision, and recall. The model demonstrated high reliability and accuracy in detecting safety helmet use, achieving an mAP of 89.5%, precision of 89.6%, and recall of 83.8%. Consequently, construction site managers are provided with a valuable tool to continuously monitoring and assessing safety compliance. This facilitates prompt intervention when workers are not wearing safety helmets. Moreover, automating the detection of safety violations, this technology offers the potential to significantly enhance the efficiency of safety management. Ultimately, this contributes to reducing workplace accidents and promoting a proactive safety culture on construction sites.

This study has successfully demonstrated the effectiveness of AI-powered technology in improving safety management on construction sites. By automating the process of safety helmet detection, this system offers several valuable benefits to various stakeholders:

- For construction site managers, the system provides a reliable and efficient tool to monitor safety compliance in real-time, enabling prompt intervention and reducing the risk of head injuries among workers.

- For site commanders and labor managers, the system offers accurate and objective data on safety helmet usage, facilitating informed decision-making and the development of targeted safety interventions.

- For workers, the presence of an automated monitoring system reinforces the importance of safety compliance and encourages a proactive approach to safety practices.

By integrating this technology into construction site safety management activities, the industry can move towards a more proactive and preventative approach to safety, ultimately contributing to a significant reduction in workplace accidents and fatalities.

Future research will focus on evaluating the system's performance in real-world construction environments, where factors such as lighting variability, worker movement, and environmental conditions may impact detection accuracy. Expanding the model to detect additional personal protective equipment, including safety vests and gloves, will enhance safety compliance monitoring. Integrating AI-based site monitoring tools, such as worker behavior analysis and risk prediction models, will further contribute to accident prevention. Additionally, efforts will be made to develop energy-efficient implementations for deployment on low-power edge devices to support broader accessibility in the construction industry.

References

[1]. K. Koc, Ö. Ekmekcioğlu, A.P. Gurgun. (2022). Accident prediction in construction using hybrid wavelet-machine learning. *Automation in Construction*, 133, 103987.

[2]. B. Shao, Z. Hu, Q. Liu, S. Chen, W. He. (2019). Fatal accident patterns of building construction activities in China. *Safety Science*, 111, 253-263.

[3]. Ministry of Labor - Invalids and Social Affairs. (2023). Announcement 4162/TB-BLDTBXH regarding the situation of occupational accidents in the first 6 months of 2023. (in Vietnam)

[4]. Ministry of Labor - Invalids and Social Affairs. (2024). Announcement 4223/TB-BLDTBXH regarding the situation of occupational accidents in the first 6 months of 2024. (in Vietnam)

[5]. J. Shen, X. Xiong, Y. Li, W. He, P. Li, X. Zheng. (2021). Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning. *Computer-Aided Civil and Infrastructure Engineering*, 36(2), 180-196.

[6]. Government. (2021). Decree No. 06/2021/ND-CP on quality management, construction and maintenance of construction works. (in Vietnam)

[7]. Government. (2022). Decree No. 16/2022/ND-CP imposing penalties for administrative violations in construction. (in Vietnam)

[8]. Z. Gao, A. Sultan. (2023). AI for Improving Construction Safety: A Systematic Literature Review. *Proceedings of 59th Annual Associated Schools of Construction*

International Conference, 4, 354-362.

[9]. S. Sakhakarmi, J. Park, C. Cho. (2019). Enhanced Machine Learning Classification Accuracy for Scaffolding Safety Using Increased Features. *Journal of Construction Engineering and Management*, 145(2), 04018133.

[10]. S. Tang, M. Golparvar-Fard. (2021). Machine Learning-Based Risk Analysis for Construction Worker Safety from Ubiquitous Site Photos and Videos. *Journal of Computing in Civil Engineering*, 35(6), 04021020.

[11]. K. Koc, Ö. Ekmekcioğlu, A.P. Gurgun. (2021). Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability status of construction workers. *Automation in Construction*, 131, 103896.

[12]. A.B.K. Rabbi, I. Jeelani. (2024). AI integration in construction safety: Current state, challenges, and future opportunities in text, vision, and audio based applications. *Automation in Construction*, 164, 105443.

[13]. G.F. Bigham, S. Adamtey, L. Onsarigo, N. Jha. (2019). Artificial Intelligence for Construction Safety: Mitigation of the Risk of Fall. In: Arai, K., Kapoor, S., and Bhatia, R. (eds.) Intelligent Systems and Applications. pp. 1024-1037. *Springer International Publishing, Cham (2019)*.

[14]. J. Choi, B. Gu, S. Chin, J.-S. Lee,. (2020). Machine learning predictive model based on national data for fatal accidents of construction workers. *Automation in Construction*, 110, 102974.

[15]. H. Baker, M.R. Hallowell, A.J.-P. Tixier. (2020). AI-based prediction of independent construction safety outcomes from universal attributes. *Automation in Construction*, 118, 103146.

[16]. H. Abbasianjahromi, M. Aghakarimi. (2021). Safety performance prediction and modification strategies for construction projects via machine learning techniques. *Engineering, Construction and Architectural Management*, 30(3), 1146-1164.

[17]. H. Wu, J. Zhao. (2018). An intelligent vision-based approach for helmet identification for work safety. *Computers in Industry*, 100, 267-277.

[18]. A.H.M. Rubaiyat, T.T. Toma, M. Kalantari-Khandani, S.A. Rahman, L. Chen, Y. Ye, C.S. Pan. (2016). Automatic Detection of Helmet Uses for Construction Safety. *2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW)*, pp. 135–142.

[19]. R. Hussain, A. Sabir, D.-Y. Lee, S.F.A. Zaidi, A. Pedro, M.S. Abbas, C. Park. (2024). Conversational AI-based VR system to improve construction safety training of migrant workers. *Automation in Construction*, 160, 105315.

[20]. È. Quintana Aguasca. (2021). Tracking by detection using YOLO and Kalman filtering.

[21]. H. Li, L. Deng, C. Yang, J. Liu, Z. Gu. (2021). Enhanced YOLO v3 Tiny Network for Real-Time Ship Detection From Visual Image. *IEEE Access*, 9, 16692-16706.

[22]. A. Hayat, F. Morgado-Dias. (2022). Deep Learning-Based Automatic Safety Helmet Detection System for Construction Safety. *Applied Sciences*, 12(16), 8268.

[23]. J. Han, Y. Liu, Z. Li, Y. Liu, B. Zhan. (2023). Safety Helmet Detection Based on YOLOv5 Driven by Super-Resolution Reconstruction. *Sensors*, 23(4), 1822.

[24]. Q. An, Y. Xu, J. Yu, M. Tang, T. Liu, F. Xu. (2023). Research on Safety Helmet Detection Algorithm Based on Improved YOLOv5s. *Sensors*. 23(13), 5824.

[25]. Kisaezehra, M.U. Farooq, M.A. Bhutto, A.K. Kazi. (2023). Real-Time Safety Helmet Detection Using Yolov5 at Construction Sites. *Intelligent Automation & Soft Computing*, 36(1), 911-927.