Journal of Advanced Sciences and Mathematics Education
Volume 6, Issue 3, 643-663.
e_ISSN: 2798-1606
DOI: 10.58524/jasme.v6i3.1462

An occlusion-aware framework for vision-based PPE detection using YOLO11x and bytetrack

Mochammad Akmal Humami*
Universitas Negeri Surabaya, INDONESIA

Monica Cinthya
Universitas Negeri Surabaya, INDONESIA

Article Info

Article history:

Received: April 27, 2026
Revised: July 10, 2026
Accepted: Sept 03, 2026

Keywords:

Construction Safety;
Multi-Object Tracking;
Personal Protective Equipment;
Yolo11x.

Abstract

Background: Ensuring workers' compliance with Personal Protective Equipment (PPE) remains challenging in construction environments because frequent worker movement and occlusion reduce the effectiveness of vision-based monitoring systems. Although recent studies have achieved promising object detection performance, limited attention has been given to the integrated evaluation of object detection and multi-object tracking under varying occlusion conditions, resulting in insufficient understanding of worker identity preservation during continuous monitoring.
Aim: This study aims to provide an integrated evaluation of object detection and multi-object tracking for vision-based PPE monitoring by combining the YOLO11x object detection model with the ByteTrack tracking algorithm. The study investigates how different occlusion conditions influence both detection accuracy and identity preservation, providing a more comprehensive assessment of integrated PPE monitoring performance.
Methods: The proposed framework was trained and evaluated using the publicly available PPE Detection Dataset, which provides annotated images of construction workers wearing different PPE categories. The framework integrates YOLO11x for object detection and ByteTrack for multi-object tracking, and its performance was evaluated under four predefined occlusion scenarios (no occlusion, partial occlusion, overlapping occlusion, and heavy occlusion) to investigate the influence of occlusion on detection and tracking performance.
Result: The proposed framework achieved an overall Precision of 0.830, Recall of 0.790, mAP@50 of 0.838, and mAP@50–95 of 0.490 for PPE object detection. The highest tracking performance was obtained under the no occlusion scenario (MOTA = 0.7262, IDF1 = 0.7271, HOTA = 0.5503), whereas tracking performance progressively declined as the level of occlusion increased, with heavy occlusion producing the largest number of identity switches. These results indicate that occlusion had a greater impact on identity preservation than on object detection performance.
Conclusion: The findings demonstrate the feasibility of integrating YOLO11x and ByteTrack for simultaneous PPE detection and worker tracking under varying occlusion conditions. The study further highlights that evaluating object detection together with multi-object tracking provides a more comprehensive assessment of vision-based PPE monitoring systems, particularly by revealing the greater influence of occlusion on identity preservation than on object detection performance.

To cite this article: Humami, M. A. & Cinthya, M. (2026). An occlusion-aware framework for vision-based PPE detection using YOLO11x and bytetrack. *Journal of Advanced Sciences and Mathematics Education, 6*(3), 643-663.

INTRODUCTION

Automated monitoring of personal protective equipment (PPE) compliance has become increasingly important for improving workplace safety in industrial and construction environments, where conventional manual inspection is labor-intensive, prone to human error, and difficult to perform continuously in dynamic worksites. Recent advances in deep learning have enabled automated vision-based PPE monitoring systems capable of improving the efficiency and consistency

***Corresponding author:**
Mochammad Akmal Humami, Universitas Negeri Surabaya, Indonesia
mochakmal.23206@mhs.unesa.ac.id

of safety inspections. However, maintaining reliable PPE detection under worker occlusion remains a significant challenge in complex construction environments, often leading to missed detections and reduced monitoring accuracy. Consequently, developing occlusion-aware vision-based PPE monitoring systems has become an important research direction for enhancing workplace safety.

Although recent advances in artificial intelligence (AI) and computer vision have significantly improved automated vision-based PPE monitoring, existing approaches still face substantial challenges in real-world construction environments. Detection performance often degrades under partial occlusion, object overlap, varying illumination, and complex backgrounds, resulting in missed detections and reduced reliability of safety compliance monitoring. These limitations restrict the practical deployment of current PPE monitoring systems and highlight the need for more robust vision-based approaches capable of maintaining reliable detection under challenging workplace conditions (Ahmad & Rahimi, 2025; Lee et al., 2023).

Recent advances in deep learning have enabled significant improvements in real-time object detection, making computer vision an increasingly effective approach for automated PPE detection in workplace environments. Among various object detection models, the YOLO family has become one of the most widely adopted approaches because of its favorable balance between detection accuracy and computational efficiency. Recent studies have reported promising performance of YOLO11-based models for recognizing multiple PPE categories in workplace environments. However, existing studies have also reported several limitations, including reduced detection performance for small or partially occluded PPE objects, limited validation on diverse real-world scenarios, and the need for further investigation using real-time video analysis. These unresolved issues indicate that achieving robust and reliable vision-based PPE detection in dynamic workplace environments remains a significant research challenge (Ordrick et al., 2025).

While object detection identifies the presence of PPE in individual frames, continuous workplace monitoring also requires maintaining worker identities across video sequences. This capability is provided by multi-object tracking (MOT), which associates detected objects across consecutive frames to support continuous monitoring of worker movements and PPE compliance. Among existing MOT algorithms, ByteTrack has demonstrated competitive tracking performance by associating both high- and low-confidence detections, thereby improving tracking continuity and reducing identity fragmentation in dynamic environments (Zhang et al., 2022). Accordingly, integrating advanced object detection and multi-object tracking provides a suitable basis for investigating occlusion-aware vision-based PPE monitoring in dynamic workplace environments.

Despite the remarkable progress in object detection and multi-object tracking, occlusion remains one of the most challenging issues in vision-based PPE monitoring. In real workplace environments, workers are frequently partially or completely obscured by machinery, construction materials, or other workers, causing missed detections, identity switches, and tracking interruptions. These conditions reduce the reliability of automated safety monitoring systems, particularly in crowded and dynamic workplaces. Although previous studies have recognized the adverse impact of occlusion on detection and tracking performance, existing detection and tracking methods still struggle to maintain reliable performance in complex scenes because occlusion frequently interrupts object visibility, disrupts identity continuity, and increases identity switches. Furthermore, most studies emphasize improvements in detection or tracking performance without systematically evaluating robustness under different levels of occlusion. This limitation highlights the need for a more comprehensive investigation of occlusion-aware PPE monitoring systems in real-world workplace environments (Luiten et al., 2021; Y. H. Wang et al., 2024).

Although previous studies have demonstrated promising results in PPE detection and multi-object tracking, several limitations remain. Previous studies have predominantly focused on improving either object detection accuracy or tracking performance, while comprehensive

evaluations of integrated detection–tracking frameworks for PPE monitoring remain limited (Lee et al., 2023; Ordrick et al., 2025; Zhang et al., 2022). Furthermore, existing studies rarely perform systematic evaluations under different levels of occlusion that reflect real workplace conditions, making it difficult to fully understand the robustness of integrated vision-based PPE monitoring systems. These limitations reveal an important research gap in evaluating the performance of integrated detection–tracking frameworks under realistic occlusion conditions.

To address this research gap, this study proposes an occlusion-aware framework for vision-based PPE detection that integrates YOLO11x and ByteTrack to support continuous PPE monitoring in dynamic workplace environments. Unlike previous studies, which primarily focused on optimizing detection or tracking performance independently, this study extends existing research by systematically evaluating the robustness of an integrated detection–tracking framework across multiple occlusion scenarios representative of real workplace conditions. This systematic evaluation provides a comprehensive assessment of integrated detection–tracking performance rather than assessing object detection and tracking separately. The findings are expected to advance the understanding of how varying levels of occlusion affect integrated detection and tracking performance, thereby providing a more rigorous basis for evaluating vision-based PPE monitoring systems and supporting the development of more reliable occupational safety monitoring systems.

LITERATURE REVIEW

Personal Protective Equipment (PPE) monitoring represents an important component of occupational safety management, particularly in construction environments where workers are continuously exposed to dynamic and potentially hazardous conditions. Conventional PPE inspection generally depends on direct human observation, which can become inefficient when monitoring large work areas, multiple workers, and continuously changing activities. Vision-based monitoring provides an alternative approach by using image and video data to automatically identify workers and determine the presence or absence of required protective equipment. The development of deep learning has further strengthened this approach by enabling computer vision systems to recognize PPE objects with greater accuracy across complex visual environments. Typical PPE monitoring tasks involve detecting safety equipment such as helmets, vests, and shoes as well as identifying corresponding non-compliance categories (Ahmed et al., 2023; Al-Azani et al., 2024; Delhi et al., 2020; Rahman et al., 2025, 2026; Xiong & Tang, 2021). Such automated recognition can support continuous safety assessment without requiring inspectors to observe every worker directly. Nevertheless, construction scenes present considerable visual complexity because workers frequently move, overlap with surrounding objects, and appear at different scales and viewing angles. Variations in illumination, background complexity, and object visibility can further reduce the reliability of PPE recognition. These challenges indicate that successful PPE monitoring depends not only on identifying protective equipment in isolated images but also on maintaining reliable observations across dynamic video sequences. Therefore, vision-based PPE monitoring should be understood as an integrated visual-analysis problem involving both accurate object recognition and continuous worker observation (Ahmad & Rahimi, 2025; Lee et al., 2023).

Object detection constitutes the primary computational component of vision-based PPE monitoring because it enables protective equipment and workers to be simultaneously localized and classified within visual scenes. The You Only Look Once (YOLO) family has become particularly relevant to this task because its single-stage architecture supports an effective balance between detection accuracy and computational efficiency. YOLO performs object localization and classification within an integrated detection pipeline, making the architecture suitable for workplace

monitoring that requires rapid processing of image and video data (Ali & Zhang, 2024; Sapkota et al., 2025; Vourkos et al., 2025; N. Wang et al., 2024). Recent developments in YOLO-based models have progressively enhanced feature extraction and object localization capabilities across increasingly complex visual conditions. Within PPE applications, YOLO-based detection can distinguish compliant and non-compliant equipment categories while simultaneously identifying workers within construction scenes. The YOLO11 generation provides a contemporary detection architecture that can accommodate multiple PPE classes within a unified vision-based framework. Nevertheless, strong aggregate detection performance does not necessarily indicate equivalent performance across individual PPE categories because differences in object scale, visibility, and visual characteristics may influence detection accuracy. Small or partially obscured PPE components remain particularly challenging because limited visual information can reduce localization and classification confidence. Consequently, multiple complementary evaluation measures are required to determine whether the detector maintains adequate recognition performance across different PPE categories and visibility conditions. Precision, Recall, mAP@50, and mAP@50–95 therefore provide relevant quantitative measures for evaluating YOLO11x as the detection component of an integrated PPE monitoring framework (Ordrick et al., 2025; Ultralytics, 2024).

Object detection alone provides information about objects appearing within individual frames but does not inherently preserve the identity of the same worker throughout a video sequence. Continuous safety monitoring therefore requires multi-object tracking (MOT), which establishes temporal associations among detected objects across consecutive frames. In construction environments, this capability enables monitoring systems to distinguish individual workers while they move through dynamic workplace scenes (Jeelani et al., 2021; J. Li et al., 2024; Patel et al., 2022; Tezcan et al., 2025). Reliable tracking becomes particularly important when PPE compliance needs to be associated with specific workers rather than interpreted as a series of independent detections. MOT systems generally combine detection information with data-association mechanisms to determine whether observations appearing in different frames correspond to the same object. Tracking performance consequently depends on both the quality of object detections and the ability of the association process to preserve identities over time. Missing or inaccurate detections can interrupt trajectories and increase the probability that a worker will be assigned a different identity after reappearing. This relationship demonstrates that detection and tracking should not be treated as entirely independent components within continuous PPE monitoring systems. Evaluating worker tracking therefore requires metrics representing overall tracking accuracy, association quality, identity preservation, and identity-switching errors. MOTA, HOTA, IDF1, and ID Switch collectively provide complementary measures for assessing these dimensions and determining whether a monitoring framework can maintain reliable worker trajectories throughout video sequences (Bernardin & Stiefelhagen, 2008; Luiten et al., 2021).

ByteTrack provides a multi-object tracking strategy designed to improve trajectory continuity by incorporating both high-confidence and low-confidence object detections during data association (Zhang et al., 2022). Conventional tracking processes may discard detections below a predetermined confidence threshold even though some lower-confidence observations can still contain useful information about objects remaining within the monitored scene. ByteTrack addresses this limitation by retaining these detections and attempting to associate them with existing trajectories rather than immediately excluding them from the tracking process (Zhang et al., 2022). This strategy is particularly relevant in construction monitoring because worker visibility can fluctuate when individuals move behind equipment, construction materials, structural elements, or other workers. When visibility decreases, detection confidence may decline even though the worker has not actually left the monitored area. Incorporating lower-confidence detections can therefore support tracking continuity and reduce unnecessary trajectory fragmentation (Zhang et al., 2022).

Integrating ByteTrack with YOLO11x establishes a sequential computational process in which YOLO11x generates worker and PPE detections while ByteTrack associates detected workers across consecutive frames (Liu et al., 2026). However, the effectiveness of this integration remains dependent on the availability of sufficiently reliable visual observations during the tracking process. Prolonged visibility loss can substantially weaken or eliminate detection evidence, thereby limiting the capacity of data association to preserve an existing worker identity. Thus, ByteTrack provides an appropriate tracking component for integrated PPE monitoring, while its performance under increasing worker occlusion requires systematic evaluation to determine the robustness and limitations of identity preservation (Zhang et al., 2022).

Occlusion constitutes a critical challenge for integrated detection–tracking systems because it simultaneously reduces object visibility and disrupts the temporal information required for reliable identity association. In construction environments, occlusion can occur when workers are obscured by machinery, construction materials, structural components, or other workers. Partial occlusion may leave sufficient visual information for continued detection, whereas prolonged or heavy occlusion can cause workers or PPE components to disappear from the detector across consecutive frames. Such interruptions create a substantial tracking challenge because the system must determine whether a subsequently visible worker corresponds to a previous trajectory or represents a different individual. Increasing occlusion can consequently generate missed detections, fragmented trajectories, incorrect associations, and identity switches even when the object detector maintains acceptable aggregate performance (Guan et al., 2025; M. Li et al., 2025; Meng et al., 2025; Peker & Altılar, 2026; S. Wang et al., 2026). Occlusion-aware evaluation is therefore necessary to distinguish the effect of reduced visibility on frame-level object detection from its influence on temporal identity preservation. Previous research has demonstrated the potential of deep-learning-based PPE detection and multi-object tracking, but evaluating either component independently provides only a partial representation of continuous monitoring reliability (Lee et al., 2023; Zhang et al., 2022). A comprehensive framework should consequently examine detection and tracking metrics jointly across systematically differentiated visibility conditions. Evaluating no occlusion, partial occlusion, overlapping occlusion, and heavy occlusion conditions provides a structured basis for determining how increasing visual obstruction affects both PPE recognition and worker identity continuity. Accordingly, an occlusion-aware YOLO11x–ByteTrack framework can provide a more comprehensive basis for evaluating vision-based PPE monitoring by integrating object detection accuracy and worker identity preservation within a unified computational assessment (Lee et al., 2023; Luiten et al., 2021; Wang et al., 2024; Zhang et al., 2022).

METHOD

Research Design

This study adopted an experimental framework evaluation design to systematically assess the performance of an occlusion-aware vision-based PPE detection framework under multiple occlusion scenarios. The research design was selected because it enables objective evaluation of framework robustness under controlled testing conditions, allowing the influence of different occlusion levels on integrated detection and tracking performance to be systematically investigated. The evaluation was structured to compare framework performance across predefined occlusion conditions using consistent detection and tracking metrics, thereby ensuring comparability across experimental scenarios. This design also enabled the identification of performance degradation patterns associated with increasing occlusion severity and provided empirical evidence regarding the framework's reliability under challenging visual conditions.

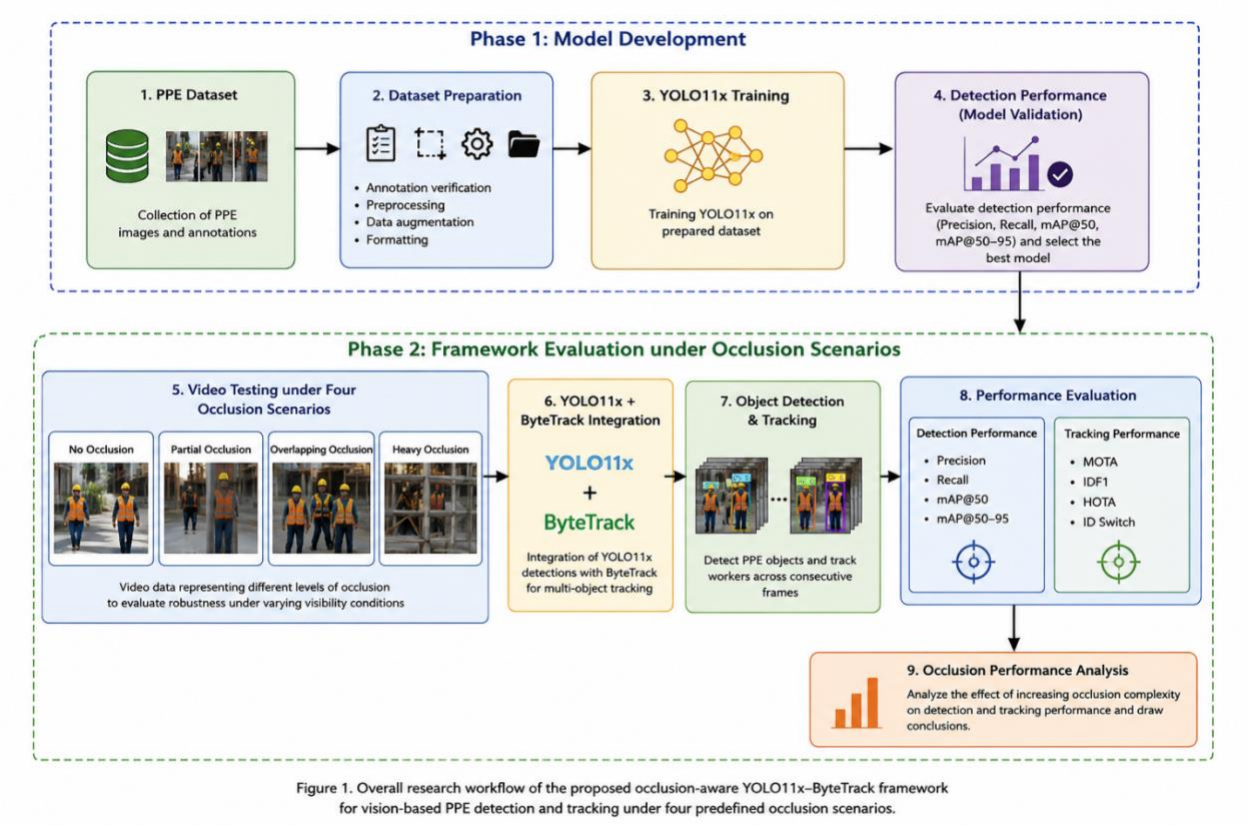


Figure 1. Research workflow of the proposed occlusion-aware YOLO11x–ByteTrack framework consisting of model development and framework evaluation under four predefined occlusion scenarios.

As illustrated in Figure 1, the research workflow consists of two main phases: model development and framework evaluation under occlusion scenarios. The model development phase focuses on dataset preparation, YOLO11x training, and detection performance validation to obtain a reliable object detection model. Subsequently, the framework evaluation phase assesses the integrated YOLO11x–ByteTrack framework using video data representing four predefined occlusion scenarios: no occlusion, partial occlusion, overlapping occlusion, and heavy occlusion. This phase systematically investigates the robustness of the integrated detection–tracking framework by jointly evaluating object detection and tracking performance under increasing levels of occlusion. The implementation details of each stage presented in Figure 1 are described in the following subsections, including dataset preparation, model training, occlusion scenario configuration, ByteTrack integration, and performance evaluation.

Dataset

This study utilized the PPE Detection Dataset obtained from Roboflow Universe, a publicly available repository for computer vision datasets. The dataset consists of annotated images of workers wearing and not wearing Personal Protective Equipment (PPE) under various workplace conditions. It includes diverse object scales, viewing angles, lighting conditions, and partial occlusions, providing representative visual variations for training and evaluating vision-based PPE detection models. Although this study employed a single public dataset, its diversity in scene characteristics and occlusion conditions makes it suitable for systematically evaluating the proposed framework under controlled experimental settings. Nevertheless, the use of a single dataset may limit the generalizability of the findings to other workplace environments, and further validation using additional datasets is recommended in future research.

The original dataset contained 4,080 annotated images distributed across 10 object classes: helmet, vest, gloves, glasses, shoes, no-helmet, no-vest, no-gloves, no-glasses, and no-shoes. To align

the dataset with the objectives of this study, only six PPE-related classes (helmet, no-helmet, vest, no-vest, shoes, and no-shoes) were retained, while the glasses, gloves, no-glasses, and no-gloves classes were excluded. In addition, the person class was automatically annotated using a pretrained YOLOv8n model trained on the COCO dataset. The generated person annotations were subsequently verified through manual visual inspection to ensure correct class assignment, accurate bounding box localization, and annotation consistency before being incorporated into the final dataset. Images that no longer contained relevant objects after the annotation adjustment process were excluded from the dataset to maintain annotation quality and dataset consistency..

After the annotation refinement process, the final dataset consisted of 3,472 annotated images representing seven object classes: person, helmet, no-helmet, vest, no-vest, shoes, and no-shoes. The dataset retained the original training, validation, and testing directory structure provided by Roboflow. Following the removal of irrelevant classes and images without valid annotations, the final distribution comprised 3,101 training images (89.31%), 254 validation images (7.32%), and 117 testing images (3.37%). The reduction in the number of samples across all subsets resulted from the annotation refinement process rather than from modifications to the original dataset partitioning. Although the testing subset contains a relatively limited number of images, the proposed framework was further evaluated using video data under four predefined occlusion scenarios, providing complementary evidence for assessing detection and tracking robustness in realistic workplace environments.

Table 1. Distribution of the PPE Detection Dataset

Dataset Subset	Number of Images	Percentage
Training Set	3,101	89.31%
Validation Set	254	7.32%
Testing Set	117	3.37%
Total	**3.472**	**100%**

The testing subset was subsequently used to evaluate the proposed YOLO11x–ByteTrack framework under four occlusion scenarios, namely No Occlusion, Partial Occlusion, Overlapping, and Heavy Occlusion.

Research Procedure

Figure 2 illustrates the implementation pipeline of the proposed YOLO11x–ByteTrack framework during the experimental stage. Unlike the overall research workflow presented in Figure 1, this pipeline describes the operational sequence used to process video data for automated PPE detection and worker tracking. The implementation begins with video input, where each video sequence is decomposed into individual frames for frame-by-frame analysis. Each extracted frame is processed by the trained YOLO11x model to detect PPE-related objects and generate corresponding bounding boxes, class labels, and confidence scores. The resulting detections are subsequently processed by the ByteTrack algorithm to associate detected workers across consecutive frames and assign persistent track identities. Finally, the integrated detection–tracking framework is evaluated under four predefined occlusion scenarios to assess its robustness using complementary object detection and multi-object tracking performance metrics. This sequential integration enables detection outputs from YOLO11x to serve directly as inputs for temporal identity association in ByteTrack, allowing object localization and tracking continuity to be evaluated within a unified processing pipeline. Performance across the four occlusion conditions is subsequently compared to determine how increasing visibility constraints affect detection accuracy, identity preservation, and overall tracking stability. The implementation details of each processing stage are presented in the following subsections.

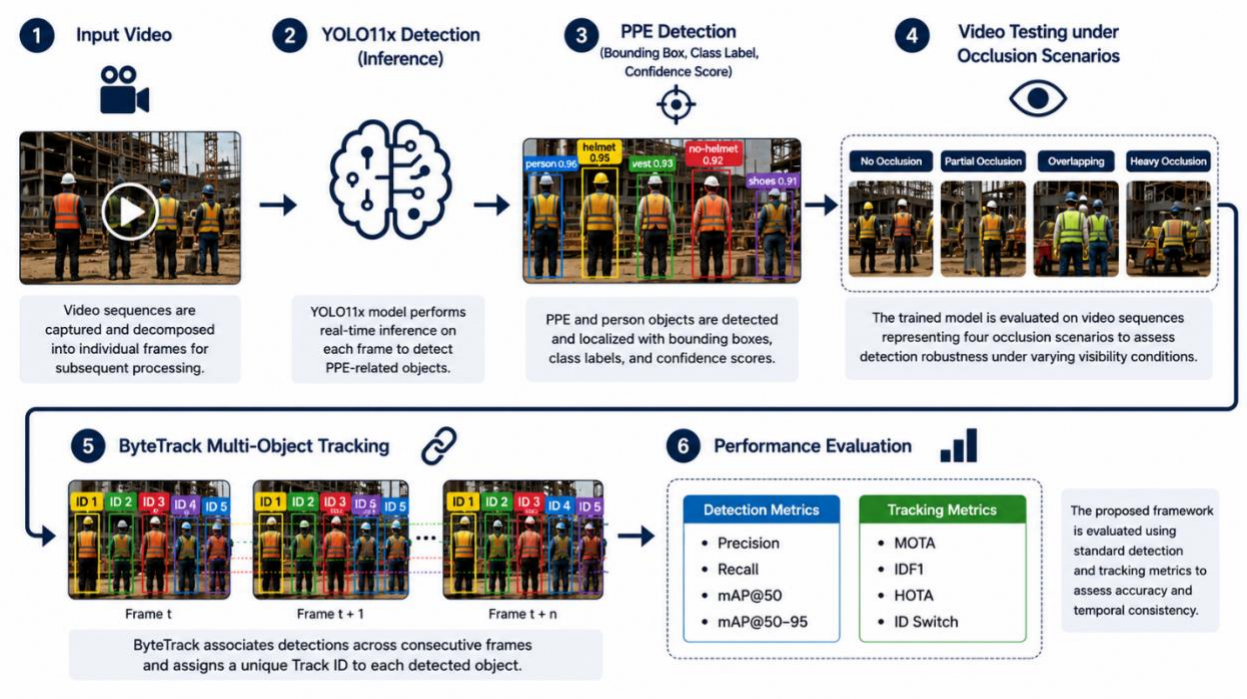


Figure 2. Processing pipeline of the proposed YOLO11x–ByteTrack framework for PPE detection and worker tracking under different occlusion scenarios.

Video data were collected from publicly available online sources containing construction workers wearing different combinations of personal protective equipment. The selected videos were manually categorized into four predefined occlusion scenarios based on visual inspection of worker visibility prior to model evaluation. The no occlusion scenario represents conditions in which the worker's body and all PPE components are fully visible. The partial occlusion scenario represents conditions in which parts of the worker's body and PPE are obscured by surrounding workers, construction materials, or equipment. The overlapping occlusion scenario represents conditions in which one worker is partially obscured by another worker, resulting in overlapping body regions while maintaining partial visibility of the PPE components. The heavy occlusion scenario represents conditions where more than 50% of the worker's body is obscured, leaving only limited visible body regions, such as the head or upper body. During inference, each extracted frame was processed using the trained YOLO11x model to detect PPE objects and generate corresponding bounding boxes, class labels, and confidence scores for each detected worker. YOLO11x was implemented using the Ultralytics framework, which performs object localization and object classification within a single-stage detection architecture (Ultralytics, 2024).

Following the object detection stage, the bounding boxes generated by YOLO11x were provided as input to the ByteTrack algorithm for multi-object tracking. ByteTrack associated detected workers across consecutive video frames and assigned persistent track identities, enabling continuous worker monitoring throughout the video sequence. In this study, ByteTrack was implemented using the default tracking configuration provided by the Ultralytics framework without modification of the tracking parameters (Ultralytics, 2024). The resulting detection and tracking outputs were subsequently evaluated to assess the robustness of the proposed framework under different occlusion scenarios. Detailed descriptions of the object detection and multi-object tracking evaluation metrics are presented in the Evaluation Metrics subsection.

Table 2. Definition of Occlusion Scenarios

Scenario	Description
No Occlusion	The worker is fully visible throughout the video sequence without any obstruction from other workers, objects, or construction equipment.
Partial Occlusion	Less than approximately 50% of the worker's body is temporarily obscured by other workers, construction materials, or equipment, while the majority of the body remains visible.
Overlapping	Two or more workers overlap within the camera view, causing their bounding boxes to intersect or appear adjacent, which increases the difficulty of object association and identity preservation during tracking.
Heavy Occlusion	More than approximately 50% of the worker's body is obscured for a prolonged period, resulting in limited visual information for reliable object detection and multi-object tracking.

Analysis Plan

The performance of the proposed YOLO11x–ByteTrack framework was evaluated using quantitative object detection and multi-object tracking metrics. The analysis was conducted on the testing dataset under four predefined occlusion scenarios, namely no occlusion, partial occlusion, overlapping occlusion, and heavy occlusion, to systematically investigate the influence of increasing occlusion on framework performance. Comparative analysis across the four occlusion scenarios was performed by jointly examining complementary object detection and tracking metrics. This analysis enabled systematic evaluation of detection accuracy, tracking consistency, and overall framework robustness under varying visibility conditions. The evaluation metrics employed in this study are described in the following subsections.

Object detection performance was evaluated using Precision, Recall, mAP@50, and mAP@50–95, which are widely accepted evaluation metrics for object detection (Everingham et al., 2015; Padilla, Rafael; Netto, Sergio L.; da Silva, 2020). In this study, these metrics were applied to evaluate the object detection performance of the proposed YOLO11x model under four predefined occlusion scenarios by assessing its ability to accurately detect and localize PPE objects under varying levels of visibility. The combined use of these complementary metrics enables a comprehensive evaluation of detection performance and facilitates comparison of model robustness across different occlusion conditions.

Multi-object tracking performance was evaluated using Multi-Object Tracking Accuracy (MOTA), Identification F1 Score (IDF1), Higher Order Tracking Accuracy (HOTA), and Identity Switch (ID Switch), which are widely adopted evaluation metrics for multi-object tracking (Bernardin & Stiefelhagen, 2008; Luiten et al., 2021). In this study, these metrics were applied to evaluate the tracking performance of the integrated YOLO11x–ByteTrack framework under four predefined occlusion scenarios by assessing tracking accuracy, identity consistency, and tracking robustness throughout the video sequences. Together, these metrics provide a comprehensive assessment of tracking performance and facilitate comparison of tracking robustness and identity preservation across different occlusion conditions.

Table 3. Performance Evaluation Metrics

Category	Metric	Description	Expected Performance
Object Detection	Precision	Ratio of correctly detected PPE objects to all predicted objects	Higher
	Recall	Ratio of correctly detected PPE objects to all ground-truth objects	Higher
	mAP@50	Mean Average Precision at IoU = 0.50	Higher

Category	Metric	Description	Expected Performance
Multi-Object Tracking	mAP@50–95	Mean Average Precision averaged over IoU thresholds from 0.50 to 0.95	Higher
	MOTA	Overall tracking accuracy considering FP, FN, and identity errors	Higher
	IDF1	Accuracy of object identity preservation across frames	Higher
	HOTA	Joint evaluation of detection and association accuracy	Higher
	ID Switch	Number of identity changes during tracking	Lower

RESULTS AND DISCUSSION

Results

The experimental results are presented to evaluate the performance of the proposed YOLO11x–ByteTrack framework under four predefined occlusion scenarios. The evaluation is organized into two parts, namely object detection and multi-object tracking performance. Before assessing the integrated framework under different occlusion conditions, the training performance of the YOLO11x model was analyzed to verify the stability of the learning process and the convergence of the trained detector.

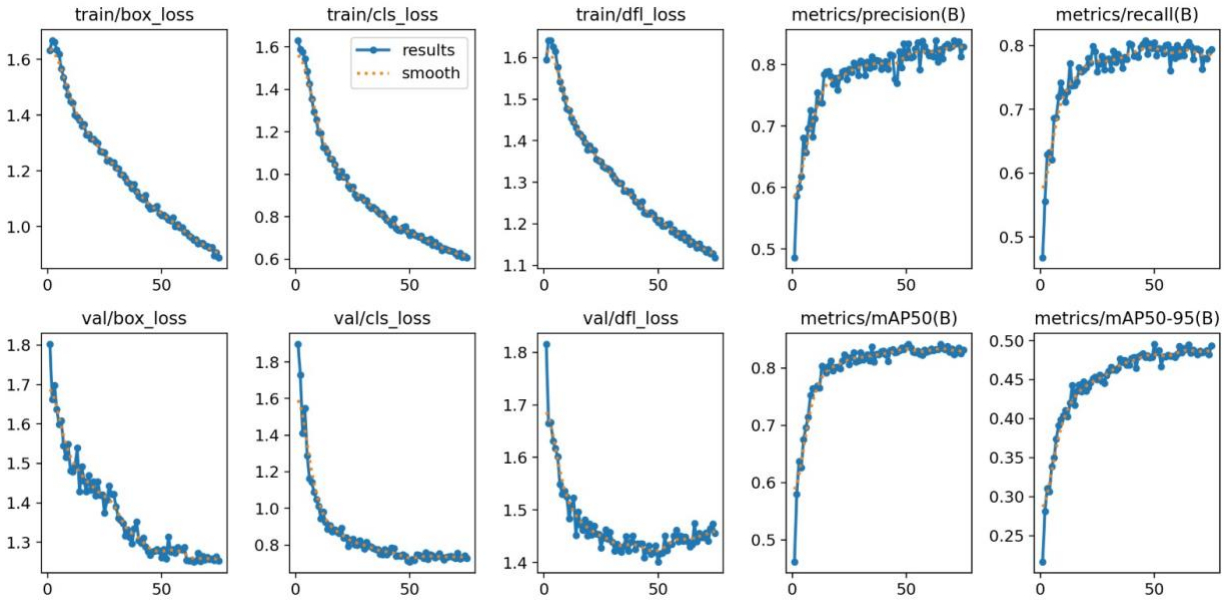


Figure 3. Training performance of the YOLO11x model showing the convergence of training and validation losses together with Precision, Recall, mAP@50, and mAP@50–95 during the training process.

Figure 3 presents the training performance of the YOLO11x model over 80 training epochs. The training and validation losses exhibited a consistent downward trend before gradually stabilizing toward the end of the training process, indicating stable convergence of the optimization procedure. Simultaneously, the evaluation metrics improved steadily throughout training, reaching a Precision of approximately 0.84, a Recall of 0.80, an mAP@50 of 0.83, and an mAP@50–95 of 0.49 at the final training epoch. These results indicate that the selected training configuration enabled the model to learn representative visual features and achieve stable convergence throughout the training

process. Consequently, the trained YOLO11x model provided a reliable basis for subsequent evaluation of the integrated YOLO11x–ByteTrack framework under different occlusion scenarios.

Table 4. Overall object detection performance of the YOLO11x model.

Metric	Value
Precision	**0.830**
Recall	**0.790**
mAP@50	**0.838**
mAP@50–95	**0.490**

Table 4 summarizes the overall object detection performance of the proposed YOLO11x model on the testing dataset. The model achieved a Precision of 0.830, Recall of 0.790, mAP@50 of 0.838, and mAP@50–95 of 0.490. These results indicate that the proposed model was able to accurately detect and localize PPE objects while maintaining a balanced trade-off between detection precision and object coverage. The overall detection performance indicates that the trained YOLO11x model provides a reliable foundation for subsequent class-wise analysis and integration with the ByteTrack algorithm for multi-object tracking.

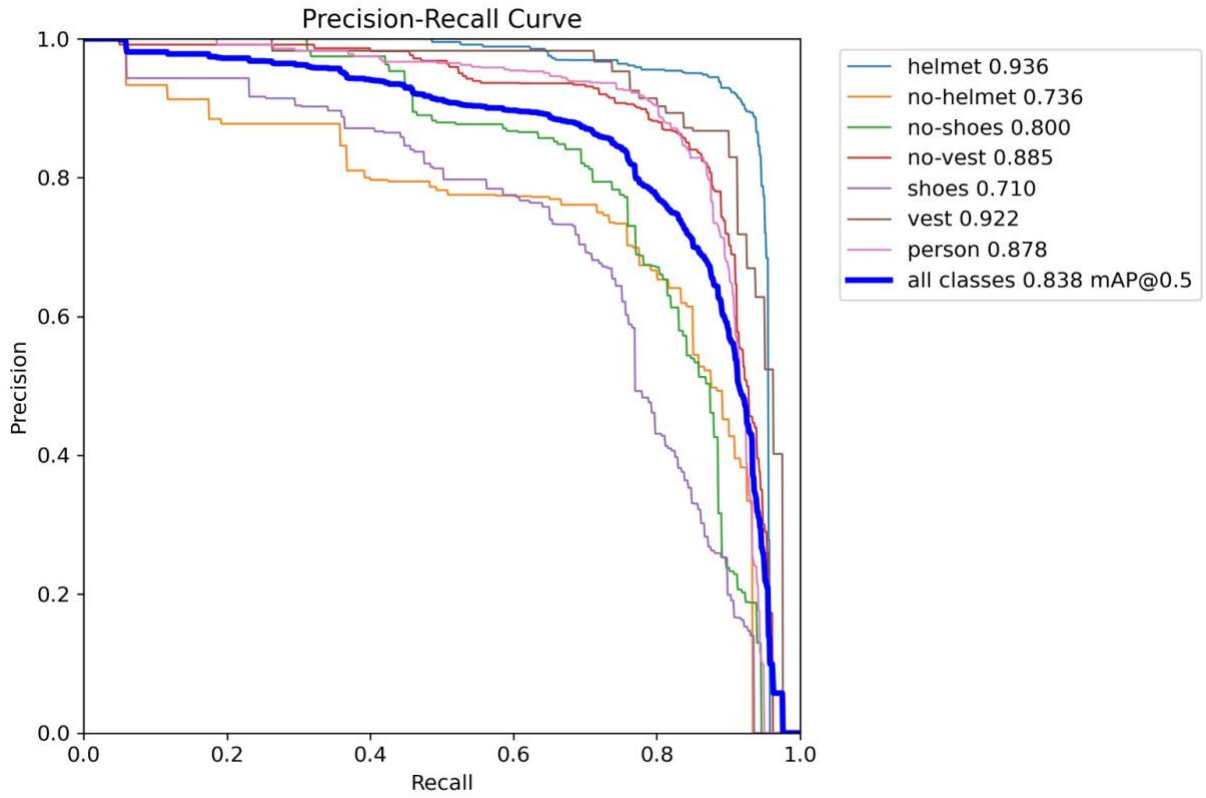


Figure 4. Precision–Recall curves of the YOLO11x model for each PPE class.

Table 5. Average Precision (AP@50) for each PPE category.

PPE Category	mAP@50
Helmet	**0.936**
Vest	**0.922**
No-Vest	**0.885**
Person	**0.878**
No-Shoes	**0.800**
No-Helmet	**0.736**

PPE Category	mAP@50
Shoes	**0.710**
Overall mAP@50	**0.838**

The Precision–Recall (PR) curves for each PPE category are presented in Figure 4, while the corresponding AP@50 values are summarized in Table 5. Overall, the PR curves indicate that the proposed YOLO11x model achieved consistent detection performance across the evaluated PPE categories, although the detection accuracy varied among individual classes. Among the evaluated classes, the helmet category achieved the highest AP@50 of 0.936, followed by the vest category with 0.922 and the no-vest category with 0.885. The higher AP@50 values obtained for the helmet and vest classes may be attributed to their distinctive visual characteristics and relatively larger visible regions, which facilitate object localization and classification. In contrast, the shoes (0.710) and no-helmet (0.736) classes achieved comparatively lower AP@50 values, reflecting the increased difficulty of distinguishing these categories under diverse workplace conditions.

Although variations in AP@50 were observed among the PPE categories, the overall Precision–Recall curve exhibited a balanced relationship between precision and recall, indicating consistent detection performance across the evaluation dataset. The overall PR curve suggests that the proposed model effectively minimized false detections while maintaining satisfactory object coverage during inference. These observations indicate that the trained YOLO11x model achieved consistent detection performance across multiple PPE categories and provides a suitable detection component for subsequent integration with the ByteTrack algorithm.

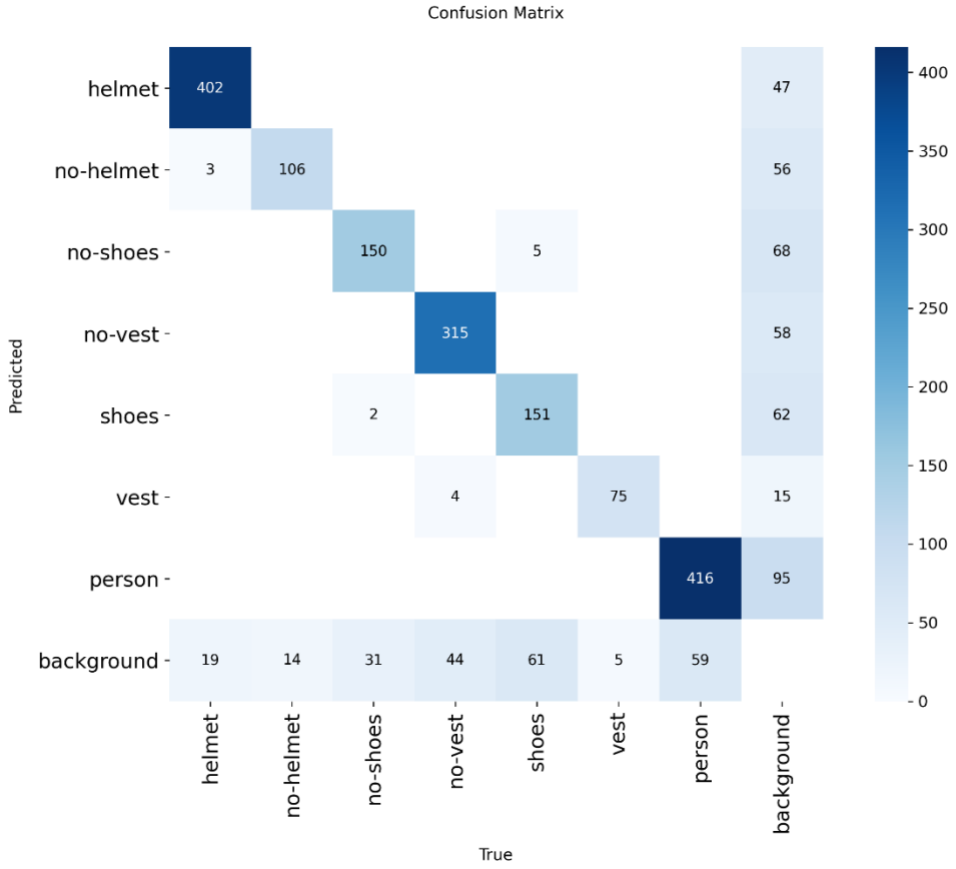


Figure 5. Confusion matrix of the YOLO11x model evaluated on the testing dataset.

Figure 5 presents the confusion matrix of the proposed YOLO11x model across all evaluated PPE categories. Most predictions are concentrated along the main diagonal, indicating that the

majority of objects were correctly classified. The helmet, person, and no-vest classes achieved the highest numbers of correct predictions, reflecting consistent classification performance on the testing dataset. The most noticeable classification errors were associated with the background class, where a number of PPE objects were predicted as background rather than being assigned to their corresponding classes. This pattern suggests that some objects were not successfully detected, which may be attributed to factors such as small object size, partial occlusion, or reduced object visibility in complex workplace scenes. In contrast, direct misclassification between visually similar PPE categories, such as helmet and no-helmet or shoes and no-shoes, occurred only infrequently. Overall, the confusion matrix indicates a high level of agreement between the predicted and ground-truth labels, suggesting consistent classification performance of the proposed YOLO11x model across the evaluated PPE categories.



Figure 6. Sample results of simultaneous PPE detection and worker identification using the proposed YOLO11x–ByteTrack framework.

Figure 6 presents a representative detection result produced by the proposed YOLO11x–ByteTrack framework on a construction-site image. Multiple workers and PPE objects were successfully localized with corresponding class labels and confidence scores, illustrating the model's capability to detect multiple objects within a complex workplace environment. The confidence scores vary across detected objects, reflecting differences in object visibility, object scale, and scene complexity. These qualitative observations are consistent with the quantitative object detection performance reported in Table 4 and the class-wise evaluation summarized in Figure 4 and Table 5. Nevertheless, several objects exhibit relatively low confidence scores or incomplete PPE detection, indicating that challenging scene conditions such as partial occlusion and viewpoint variation continue to influence detection performance. The observed variation further suggests that detection reliability depends on the amount of visible object information available to the model, particularly when PPE components are small, partially obscured, or positioned at unfavorable viewing angles. These qualitative results therefore complement the quantitative evaluation by illustrating both the practical detection capability of the proposed framework and the remaining challenges associated with maintaining consistent PPE recognition under visually complex construction-site conditions.

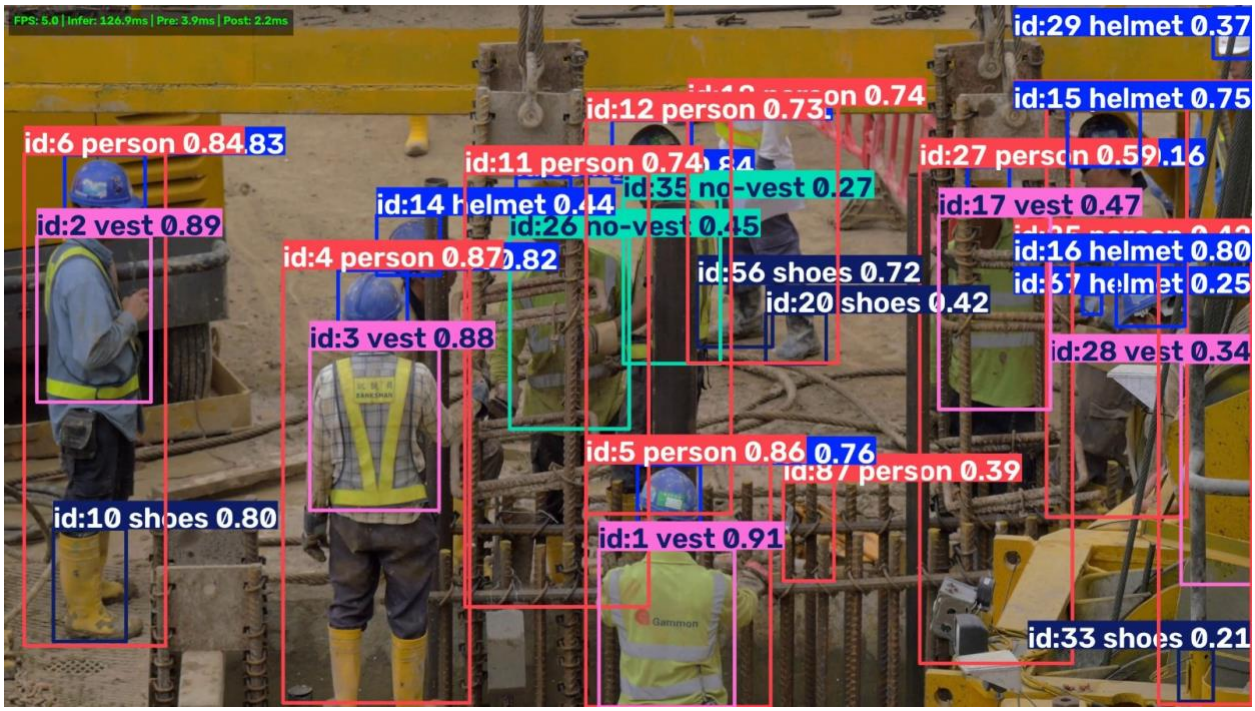


Figure 7. Continuous worker tracking using the ByteTrack algorithm.

Figure 7 presents representative multi-object tracking results produced by the proposed YOLO11x–ByteTrack framework. Each detected worker is assigned a unique track ID, illustrating the tracking process applied to multiple workers in a construction-site environment. The quantitative evaluation of tracking performance under different occlusion scenarios is presented in Table 6.

Table 6. Tracking performance under different occlusion scenarios.

Scenario	MOTA	HOTA	IDF1	ID Switch
No Occlusion	0.7262	0.5503	0.7271	5
Partial Occlusion	0.5041	0.5387	0.7158	5
Overlapping	0.4721	0.4644	0.6364	3
Heavy Occlusion	0.3388	0.4276	0.5421	41

The quantitative tracking performance under different occlusion scenarios is summarized in Table 6. The highest tracking performance was achieved under the no occlusion scenario, with a MOTA of 0.7262, HOTA of 0.5503, and IDF1 of 0.7271. As the level of occlusion increased, MOTA, HOTA, and IDF1 generally decreased, indicating a gradual reduction in tracking accuracy and identity preservation. The most significant performance degradation occurred under the heavy occlusion scenario, where MOTA and IDF1 decreased to 0.3388 and 0.5421, respectively, while the number of identity switches increased sharply from 5 in the no occlusion scenario to 41. This substantial increase suggests that prolonged object visibility loss interrupted continuous object trajectories, making data association across consecutive frames considerably more challenging. Although ByteTrack associates both high- and low-confidence detections to improve identity preservation, severe occlusion reduces the availability of reliable detections, increasing the likelihood of identity reassignment when objects reappear.

Figure 8 compares the MOTA, HOTA, and IDF1 scores obtained under the four predefined occlusion scenarios. The highest tracking performance was achieved under the no occlusion scenario, with MOTA, HOTA, and IDF1 values of 72.6%, 55.0%, and 72.7%, respectively. As the level of occlusion increased, all three metrics exhibited a progressive decline, indicating that reduced object visibility adversely affected both detection quality and object association during tracking. Among the evaluated metrics, MOTA showed the largest decrease, reflecting the accumulation of missed

detections and tracking errors under increasingly challenging conditions. In contrast, HOTA exhibited a more gradual decline, suggesting that the proposed framework retained a certain level of association performance despite the reduction in detection accuracy. Similarly, IDF1 decreased from 72.7% to 54.2%, indicating reduced identity preservation as prolonged occlusion increased the likelihood of interrupted object trajectories and identity reassignment.

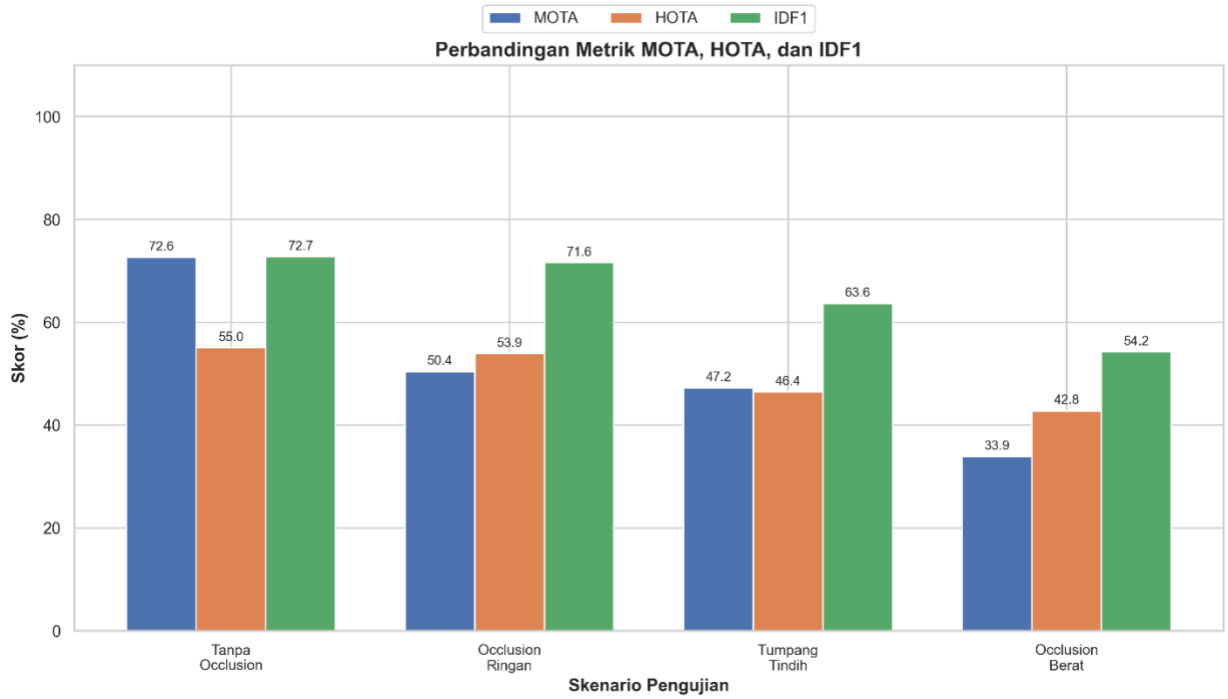


Figure 8. Comparison of MOTA, HOTA and IDF1 under different occlusion scenarios.

Figure 9 illustrates the distribution of identity switches across the four evaluated occlusion scenarios. Under the no occlusion, partial occlusion, and overlapping scenarios, only a small number of identity switches were observed, with 5, 5, and 3 identity switches, respectively, indicating that worker identities were generally preserved when object visibility remained sufficient. In contrast, the heavy occlusion scenario resulted in a substantial increase to 41 identity switches, reflecting the difficulty of maintaining continuous object identities when workers were obscured for extended periods. This increase suggests that prolonged visibility loss interrupted object trajectories and reduced the availability of reliable detections, making object association across consecutive frames more challenging. Consequently, increasing occlusion adversely affected identity preservation and overall tracking performance, with the most pronounced degradation occurring under the heavy occlusion scenario. The relatively stable identity-switch counts observed in the first three scenarios suggest that ByteTrack can maintain reliable temporal associations when sufficient detection information remains available between consecutive frames. However, under heavy occlusion, prolonged interruptions in worker visibility can weaken the continuity of trajectories and increase uncertainty during the reassociation process. The sharp increase from 3 identity switches under overlapping occlusion to 41 under heavy occlusion further indicates that occlusion duration and visibility loss may be more disruptive to identity preservation than temporary spatial overlap alone. This pattern highlights identity consistency as a particularly sensitive component of the proposed framework when operating under severe visual obstruction. Therefore, improving worker reassociation after prolonged occlusion represents an important direction for enhancing the robustness of the YOLO11x–ByteTrack framework in complex and densely populated construction environments.

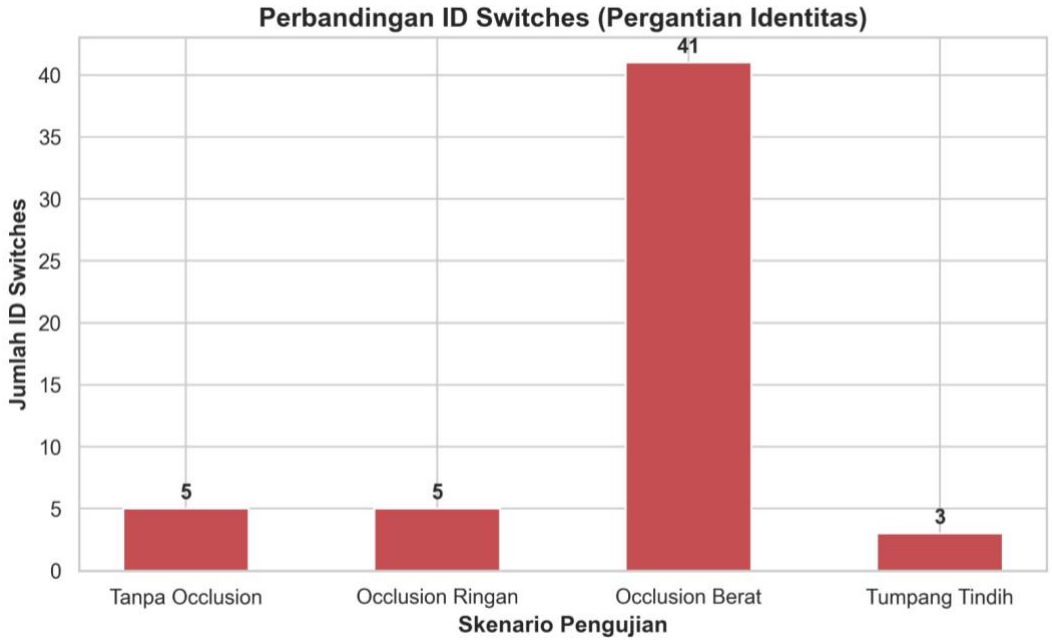


Figure 9. Comparison of identity switches under different occlusion scenarios.

Discussion

The experimental results indicate that the proposed YOLO11x–ByteTrack framework achieved consistent object detection performance while tracking performance gradually declined as the level of occlusion increased. The object detection results suggest that the trained YOLO11x model was able to accurately detect multiple PPE categories across the evaluation dataset. In contrast, the tracking results showed that identity preservation became increasingly challenging under severe occlusion, as reflected by the decreases in MOTA, HOTA, and IDF1 together with the substantial increase in identity switches. These findings suggest that integrating YOLO11x with ByteTrack is a feasible approach for continuous PPE monitoring in construction environments, although tracking performance remains sensitive to prolonged object occlusion.

The class-wise evaluation revealed noticeable differences in detection performance among the evaluated PPE categories. As summarized in Table 5, the helmet and vest categories achieved the highest AP@50 values of 0.936 and 0.922, respectively, whereas the shoes and no-helmet categories obtained comparatively lower AP@50 values of 0.710 and 0.736. These quantitative results indicate that detection performance varied across PPE categories, reflecting differences in the model's ability to accurately detect individual object classes. Unlike previous studies, which primarily focused on evaluating object detection performance, the present study further examined the influence of four predefined occlusion scenarios on the integrated YOLO11x–ByteTrack framework. The results showed that although object detection performance remained relatively consistent across the evaluated PPE categories, tracking performance deteriorated as the level of occlusion increased, highlighting the importance of jointly evaluating object detection and multi-object tracking when developing vision-based PPE monitoring systems for construction environments.

The tracking evaluation showed that the proposed YOLO11x–ByteTrack framework maintained higher tracking performance under the no occlusion and partial occlusion scenarios than under severe occlusion conditions. As summarized in Table 6, MOTA, HOTA, and IDF1 gradually decreased as the level of occlusion increased, while the number of identity switches increased substantially under the heavy occlusion scenario. These findings indicate that prolonged object visibility loss made continuous object association more difficult, resulting in interrupted trajectories and more frequent identity reassignment. Although the framework was able to preserve worker

identities under low and moderate occlusion, severe occlusion remained a major challenge because prolonged interruptions in object visibility reduced the continuity of the tracking process. This observation is consistent with previous studies reporting that prolonged occlusion remains one of the primary challenges in multi-object tracking (Zhang et al., 2022), who reported that although ByteTrack improves tracking continuity by utilizing low-confidence detections, prolonged occlusion remains a major challenge for multi-object tracking. Overall, the present results suggest that the proposed framework is suitable for construction scenarios with limited to moderate occlusion, while additional object association techniques, such as appearance-based re-identification, may further improve identity preservation under severe occlusion.

The present study extends the evaluation of vision-based PPE monitoring by jointly assessing object detection and multi-object tracking under four predefined occlusion scenarios. While previous studies primarily emphasized either PPE detection accuracy (Ordrick et al., 2025) or multi-object tracking performance (Zhang et al., 2022), and Lee et al. (2023) demonstrated the feasibility of integrating detection and tracking for construction safety monitoring, the present study systematically evaluates how different levels of occlusion influence both detection accuracy and identity preservation within a unified framework. Although the proposed framework did not achieve the highest detection or tracking performance reported in previous studies, the comparative results summarized in Table 7 indicate that it provides a broader evaluation by combining complementary detection and tracking metrics, including Precision, Recall, mAP@50, mAP@50–95, MOTA, IDF1, HOTA, and Identity Switch. More importantly, the findings reveal that increasing occlusion had a substantially greater impact on tracking performance than on object detection performance, highlighting the importance of jointly evaluating detection accuracy and identity preservation when developing vision-based PPE monitoring systems for dynamic construction environments.

Table 7. Comparison of the Proposed Framework with Previous Studies

Study	Detection Model	Tracking	Detection Performance	Tracking Performance	Occlusion Evaluation	Main Contribution
(Lee et al., 2023)	YOLO-based	DeepSORT	mAP@50: 66.4% Accuracy: 91.3%	Not reported	Limited	Integrated PPE detection and tracking.
(Zhang et al., 2022)	YOLOX	ByteTrack	Not reported	MOTA: 80.3%, IDF1: 77.3%, HOTA: 63.1%	General MOT	Robust MOT with ByteTrack.
(Ordrick et al., 2025)	YOLO11	Not reported	Precision: 94.0%, Recall: 90.1%, mAP@50: 92.8%	Not reported	Not reported	Evaluated YOLO11 for multi-class PPE detection.
This Study	**YOLO11x**	**ByteTrack**	**Precision: 83.0%, Recall: 79.0%, mAP@50: 83.8%, mAP@50–95: 49.0%**	**MOTA: 72.6%, IDF1: 72.7%, HOTA: 55.0%, IDS: 5**	**No, Partial, Overlapping, and Heavy Occlusion**	**Integrated PPE detection and MOT under occlusion**

Overall, the results indicate that the integration of YOLO11x and ByteTrack provides a feasible approach for vision-based PPE monitoring in construction environments. The proposed framework achieved an mAP@50 of 83.8% for object detection and maintained the highest tracking performance under the no occlusion scenario, with a MOTA of 72.6%, an IDF1 of 72.7%, and only five

identity switches. However, tracking performance declined substantially as the level of occlusion increased, indicating that identity preservation remains more sensitive to severe visibility loss than object detection. These findings highlight the importance of jointly evaluating object detection and multi-object tracking when developing vision-based PPE monitoring systems, particularly for environments where frequent worker occlusion is expected.

Implications

The findings of this study suggest that integrating YOLO11x and ByteTrack provides a feasible approach for vision-based PPE monitoring under varying occlusion conditions. The experimental evaluation showed that object detection performance remained relatively consistent across the evaluated PPE categories, whereas tracking performance became increasingly affected as the level of occlusion increased, particularly under heavy occlusion. These findings emphasize the importance of jointly evaluating object detection accuracy and identity preservation when assessing integrated PPE monitoring systems, as both components contribute to overall system performance in dynamic construction environments.

From a practical perspective, the proposed YOLO11x–ByteTrack framework may support the development of automated PPE monitoring systems for construction environments. By combining object detection and multi-object tracking, the framework enables simultaneous evaluation of PPE compliance and worker identity preservation within video sequences, providing more comprehensive monitoring information than object detection alone. However, because the present study was conducted under experimental conditions, additional validation in real construction environments is required to assess operational performance, computational efficiency, and long-term applicability before practical deployment.

From an academic perspective, this study contributes to the evaluation of integrated vision-based PPE monitoring systems by examining the different effects of occlusion on object detection and multi-object tracking performance. Rather than proposing a new detection or tracking algorithm, the study demonstrates that increasing occlusion has a greater impact on identity preservation than on object detection performance. This finding provides additional insight into the behavior of integrated detection–tracking frameworks under varying visibility conditions and suggests that future evaluations of PPE monitoring systems should incorporate both object detection and multi-object tracking metrics to achieve a more comprehensive assessment of system performance.

Limitations and Suggestions for Future Research

This study has several limitations that should be considered when interpreting its findings. Content validation involved only one expert, so the results represent preliminary rather than comprehensive evidence of instrument validity. The needs analysis was also based on brief informal interviews without audio recordings or verbatim transcripts, limiting the depth of qualitative interpretation. In addition, the instrument has not yet been administered to the planned sample of graduate mathematics teachers. Consequently, its construct validity, reliability, and psychometric performance remain to be established. The lower relevance score for TCK also indicates that several technology-content items require further refinement. Future research should involve multiple independent experts to strengthen content-validation evidence. The revised instrument should subsequently be tested with a larger and more diverse sample of mathematics teachers. Further psychometric analysis is needed to examine its factor structure, internal consistency, and relationships among the proposed dimensions. Recorded interviews and classroom observations could also provide deeper evidence regarding teachers' technology-integration practices. Future studies should further examine whether the TPACK and Didactic Tetrahedron dimensions remain empirically distinguishable when integrated within a single instrument. Overall, iterative expert

review, broader empirical testing, and classroom-based investigation are required before the instrument can be recommended for wider application.

CONCLUSION

The experimental results indicate that the proposed YOLO11x–ByteTrack framework achieved an overall Precision of 0.830, Recall of 0.790, mAP@50 of 0.838, and mAP@50–95 of 0.490 for PPE object detection. The evaluation further showed that tracking performance was highest under the no occlusion scenario and gradually decreased as the level of occlusion increased, with severe occlusion producing the largest number of identity switches and the lowest MOTA, IDF1, and HOTA scores. These findings indicate that increasing occlusion had a greater impact on identity preservation than on object detection performance, highlighting the importance of jointly evaluating object detection and multi-object tracking when assessing vision-based PPE monitoring systems. The study also demonstrates the feasibility of integrating YOLO11x and ByteTrack within a unified framework for evaluating PPE detection and worker tracking under varying visibility conditions. These findings provide additional evidence that evaluating object detection together with multi-object tracking offers a more comprehensive assessment of vision-based PPE monitoring systems operating under different occlusion conditions.

The results indicate that the integration of YOLO11x and ByteTrack enables simultaneous evaluation of PPE object detection and worker tracking under varying occlusion conditions. The experimental findings showed that object detection performance remained relatively consistent across the evaluated PPE categories, whereas tracking performance deteriorated more substantially as occlusion increased, indicating that identity preservation is more sensitive to visibility loss than object detection. This finding highlights the importance of incorporating both object detection and multi-object tracking metrics when evaluating integrated vision-based PPE monitoring systems. Under the experimental conditions investigated in this study, the proposed framework demonstrates the feasibility of integrating object detection and multi-object tracking for evaluating PPE monitoring performance under different levels of occlusion.

These findings demonstrate that evaluating object detection performance together with multi-object tracking provides a more comprehensive assessment of vision-based PPE monitoring systems than relying on object detection metrics alone. In particular, the results show that increasing occlusion has a greater impact on identity preservation than on object detection performance, highlighting the importance of incorporating both detection and tracking metrics when evaluating integrated PPE monitoring systems for construction environments.

AUTHOR CONTRIBUTIONS STATEMENT

Mochammad Akmal Humami contribued conceptualization, methodology, software, data curation, formal analysis, investigation, validation, visualization, writing original draft. monica cinthya contributed conceptualization, methodology, supervision, validation, project administration, writing – review & editing, corresponding author.

REFERENCES

Ahmad, H. M., & Rahimi, A. (2025). A dataset for human safety and personal protective equipment detection in manufacturing industry. *Journal of Safety Science and Resilience*, *6*(2), 175–185. https://doi.org/10.1016/j.jnlssr.2024.09.002

Ahmed, M. I. B., Saraireh, L., Rahman, A., Al-Qarawi, S., Mhran, A., Al-Jalaoud, J., Al-Mudaifer, D., Al-Haidar, F., AlKhulaifi, D., Youldash, M., & Gollapalli, M. (2023). Personal protective equipment detection: A deep-learning-based sustainable approach. *Sustainability*, *15*(18), Article 13990. https://doi.org/10.3390/su151813990

Al-Azani, S., Luqman, H., Alfarraj, M., Sidig, A. A. I., Khan, A. H., & Al-Hamed, D. (2024). Real-time monitoring of personal protective equipment compliance in surveillance cameras. *IEEE Access*, *12*, 121882–121895. https://doi.org/10.1109/ACCESS.2024.3451117

Ali, M. L., & Zhang, Z. (2024). The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection. *Computers*, *13*(12), Article 336. https://doi.org/10.3390/computers13120336

Bernardin, K., & Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: The CLEAR MOT metrics. *EURASIP Journal on Image and Video Processing*, *2008*, Article 246309. https://doi.org/10.1155/2008/246309

Delhi, V. S. K., Sankarlal, R., & Thomas, A. (2020). Detection of personal protective equipment (PPE) compliance on construction site using computer vision based deep learning techniques. *Frontiers in Built Environment*, *6*, Article 136. https://doi.org/10.3389/fbuil.2020.00136

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2015). The PASCAL Visual Object Classes Challenge: A retrospective. *International Journal of Computer Vision*, *111*(1), 98–136. https://doi.org/10.1007/s11263-014-0733-5

Guan, Z., Wang, Z., Zhang, G., Li, L., Zhang, M., Shi, Z., & Jiang, N. (2025). Multi-object tracking review: Retrospective and emerging trend. *Artificial Intelligence Review*, *58*(8), Article 235. https://doi.org/10.1007/s10462-025-11212-y

Jeelani, I., Asadi, K., Ramshankar, H., Han, K., & Albert, A. (2021). Real-time vision-based worker localization & hazard detection for construction. *Automation in Construction*, *121*, Article 103448. https://doi.org/10.1016/j.autcon.2020.103448

Lee, Y. R., Jung, S. H., Kang, K. S., Ryu, H. C., & Ryu, H. G. (2023). Deep learning-based framework for monitoring wearing personal protective equipment on construction sites. *Journal of Computational Design and Engineering*, *10*(2), 905–917. https://doi.org/10.1093/jcde/qwad019

Li, J., Miao, Q., Zou, Z., Gao, H., Zhang, L., Li, Z., & Wang, N. (2024). A review of computer vision-based monitoring approaches for construction workers' work-related behaviors. *IEEE Access*, *12*, 7134–7155. https://doi.org/10.1109/ACCESS.2024.3350773

Li, M., Zhang, Y., Jia, Y., & Yang, Y. (2025). Advancing multi-object tracking through occlusion-awareness and trajectory optimization. *Knowledge-Based Systems*, *310*, Article 112930. https://doi.org/10.1016/j.knosys.2024.112930

Liu, T., Zong, L., Jie, Y. X., Li, T., Ji Nan, S., & Jun Peng, H. (2026). YOLO-BIM-LSTM: An auditable integration framework for real-time construction site monitoring, BIM semantic linking, and cost forecasting. *IEEE Access*, *14*, 73130–73152. https://doi.org/10.1109/ACCESS.2026.3691938

Luiten, J., Ošep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., & Leibe, B. (2021). HOTA: A higher order metric for evaluating multi-object tracking. *International Journal of Computer Vision*, *129*(2), 548–578. https://doi.org/10.1007/s11263-020-01375-2

Meng, W., Duan, S., Ma, S., & Hu, B. (2025). Motion-perception multi-object tracking (MPMOT): Enhancing multi-object tracking performance via motion-aware data association and trajectory connection. *Journal of Imaging*, *11*(5), Article 144. https://doi.org/10.3390/jimaging11050144

Ordrick, J., Wibowo, G. H., Fahmi, A., Kurniawan, I., & Haq, E. S. (2025). YOLOv11-based automated PPE detection system for workplace safety monitoring in electric power distribution operations. *Journal of Information Systems and Informatics*, *7*(4), 4294–4329. https://doi.org/10.63158/journalisi.v7i4.1379

Padilla, R., Netto, S. L., & da Silva, E. A. B. (2020). A survey on performance metrics for object-detection algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)* (pp. 237–242). IEEE. https://doi.org/10.1109/IWSSIP48289.2020.9145130

Patel, V., Chesmore, A., Legner, C. M., & Pandey, S. (2022). Trends in workplace wearable technologies and connected-worker solutions for next-generation occupational safety, health, and productivity. *Advanced Intelligent Systems*, *4*(1), Article 2100099. https://doi.org/10.1002/aisy.202100099

Peker, K., & Altılar, D. T. (2026). Occlusion handling algorithms in visual object tracking: A survey. *Applied Intelligence*, *56*(10), Article 351. https://doi.org/10.1007/s10489-026-07345-4

Rahman, A., Ahmed, M. S., AlBugami, K. N., Alabbad, A. Y., AlFantoukh, A. A., Alshaikhahmed, Y. H., Alzahrani, Z. S., Khan, M. A. A., Youldash, M., & Alshahrani, S. M. (2026). PPE-EYE: A deep learning approach to personal protective equipment compliance detection. *Computers*, *15*(1), Article 45. https://doi.org/10.3390/computers15010045

Rahman, A., Alatallah, F. A., Almubarak, A. J., Alkhazal, H. A., Alzayer, H. A., Shaaban, Y. Z., Min-Allah, N., Bakry, A., & Aloup, K. (2025). Deep learning-based automated inspection of generic personal protective equipment. *Computers, Materials & Continua*, *85*(2), 3507–3525. https://doi.org/10.32604/cmc.2025.067547

Sapkota, R., Flores-Calero, M., Qureshi, R., Badgujar, C., Nepal, U., Poulose, A., Zeno, P., Vaddevolu, U. B. P., Khan, S., Shoman, M., Yan, H., & Karkee, M. (2025). YOLO advances to its genesis: A decadal and comprehensive review of the You Only Look Once (YOLO) series. *Artificial Intelligence Review*, *58*(9), Article 274. https://doi.org/10.1007/s10462-025-11253-3

Tezcan, Ö., Akcay, C., Sari, M., & Cavus, M. (2025). Sensor-based automatic recognition of construction worker activities using deep learning network. *Sensors*, *25*(13), Article 3988. https://doi.org/10.3390/s25133988

Ultralytics. (2024). *Ultralytics YOLO documentation*. https://docs.ultralytics.com/

Vourkos, E. G., Christoforou, E. G., Toulkeridou, E., Ramdani, N., & Panayides, A. S. (2025). Real-time object and human detection using YOLO for autonomous robot navigation in healthcare workspaces. *SN Computer Science*, *6*(5), Article 545. https://doi.org/10.1007/s42979-025-04042-w

Wang, N., Duan, Y., Li, S., Huang, F., & Qiao, L. (2024). Advanced YOLO-DeepSort-based system for drainage pipeline defects intelligent detection. *IEEE Access*, *12*, 145857–145870. https://doi.org/10.1109/ACCESS.2024.3462738

Wang, S., Wang, Y., Wang, B., Wei, C., & Liu, H. (2026). Dynamic occlusion-predictive neural network for robust roadside multi-vehicle tracking. *Sensors*, *26*(11), Article 3529. https://doi.org/10.3390/s26113529

Wang, Y. H., Hsieh, J. W., Chen, P. Y., Chang, M. C., So, H. H., & Li, X. (2024). SMILEtrack: Similarity learning for occlusion-aware multiple object tracking. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(6), 5740–5748. https://doi.org/10.1609/aaai.v38i6.28386

Xiong, R., & Tang, P. (2021). Pose guided anchoring for detecting proper use of personal protective equipment. *Automation in Construction*, *130*, Article 103828. https://doi.org/10.1016/j.autcon.2021.103828

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Eds.), *Computer Vision – ECCV 2022* (pp. 1–21). Springer. https://doi.org/10.1007/978-3-031-20047-2_1