Vol. 16 (2026): Proceedings of the 2026 International Conference on AI and Technological Innovation Applications (AITIA 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/aitia.2026131139

Application of Deep Learning in Visual Recognition of Unsafe Behaviors of Construction Personnel

Feifan Peng*

Shanxi University of Finance and Economics, Management Science and Engineering, Taiyuan 030006, China

**Corresponding author: Feifan Peng.*

Abstract

Unsafe behavior of construction personnel is a primary cause of the frequent safety accidents in the construction industry. The traditional manual inspection method has some disadvantages, such as limited coverage, strong subjectivity, poor real-time performance and so on. With the development of deep learning and computer vision technology, automatic recognition of unsafe behavior based on vision has become an important research direction of construction safety management. This paper systematically reviews the progress in the application of deep learning in visual recognition of unsafe behaviors of construction workers. Firstly, it analyzes the formation mechanism of unsafe behavior and the shortcomings of traditional management methods from the perspective of behavior safety theory. Then according to the technical route, the existing methods are divided into four categories: target detection, pose estimation, time series modeling and lightweight deployment. The principles, representative models and typical application scenarios of various methods are summarized respectively, and the applicability and limitations of various methods in the actual site application are analyzed. The analysis shows that the deep learning recognition method has been widely used in static protective equipment detection, dynamic dangerous action recognition and real-time early warning, but it still faces challenges such as small target detection difficulty, lack of robustness in complex environment, and difficulty in balancing real-time performance and accuracy. Future research should focus on multimodal fusion, video understanding of large models, online continuous learning and interpretable AI, thereby promoting the intelligent upgrading of construction safety management.

Keywords

construction safety, unsafe behavior, deep learning, computer vision

1. Introduction

According to the data of the National Bureau of statistics, a total of 19626 people were killed in various production safety accidents in 2024, and the workplace safety situation remains severe. From the perspective of the composition of accident deaths, the construction industry accounts for a relatively high proportion of all kinds of accidents. The primary cause of construction safety accidents is human unsafe behavior [1]. According to relevant statistics, human accidents account for more than 96.5%, especially in the construction industry, the proportion of accidents caused by human unsafe behavior is as high as 85% [2]. Therefore, effective

identification, early warning and intervention of unsafe behaviors of construction personnel are the key factors to prevent accidents at the source.

According to the formation mechanism of unsafe behavior, the behavioral safety management meta model believes that the generation of unsafe behavior is a process from implicit safe behavior to explicit safe behavior and then to the result of safe behavior, and the root cause of the failure of behavioral safety management is the lack of organizational safety culture [3].

The traditional supervision mode of unsafe behavior mainly relies on the on-site inspection of safety officers and video review afterwards. However, there are some defects in manual inspection, such as limited coverage, easy to miss or misjudge, and unable to give real-time warning. Although the traditional behavior safety management (BBS) has been applied in the construction industry, it has many limitations, such as the intervention effect is difficult to guarantee, and the analysis of behavior generation mechanism is insufficient [2]. Therefore, an intelligent method that can automatically analyze video content and identify unsafe behaviors in real time is needed.

The rapid development of computer vision technology, especially the wide application of deep learning, provides a strong support for the intelligent construction safety management. Different from the traditional image processing methods that mainly rely on manual feature setting, the deep learning model can automatically learn the feature representation from a large number of data. It has made great progress in target detection, pose estimation, action recognition and other tasks. The practical application in Wuhan Lianghu tunnel project shows that the intelligent recognition and correction framework of unsafe behaviors based on machine vision can effectively identify common unsafe behaviors on site, and realize the whole cycle management of in the process, after the event and in advance through personalized training and trajectory prediction [4].

This paper aims to systematically review the research progress of in-depth learning in the field of visual recognition of unsafe behaviors of construction workers, and classify and summarize the technologies such as target detection, pose estimation and time series modeling. Combined with specific cases, the applicable scenarios of various methods are analyzed, in order to provide reference for relevant research and engineering application.

2. Deep Learning Technical Approaches

2.1 Method Based on Target Detection

Target detection is one of the basic tasks in the field of computer vision. It locates and recognizes the target object from the image. In recent years, with the rapid development of deep convolutional neural network, the target detection method based on deep learning is widely used in the identification of unsafe behaviors of construction personnel, such as detecting the wearing of protective equipment such as helmets and reflective clothing, or monitoring whether the personnel operate the construction machinery correctly. Based on different network structures, the target detection algorithm based on deep learning can be divided into two-stage algorithm and single-stage algorithm.

A typical two-stage algorithm is the region-based convolutional neural network (R-CNN) series, such as Faster R-CNN. Its working principle is to generate candidate regions, extract and classify candidate regions, and finally adjust them with regression model. Faster R-CNN can recognize construction personnel, excavators, tower cranes and other engineering entities from the site monitoring images, and judge the spatial relationship between entities through graph neural network (GNN), which provides the basis for unsafe behavior recognition. Moreover, it can also be applied to the monitoring of the hoisting safety distance of the tower crane, accurately identify and locate the hook and workers, and calculate the horizontal distance between them, which can effectively prevent the occurrence of accidents caused by falling objects [5]. Although the accuracy of this method is high, the detection speed is slow, and it is easy to be affected by occlusion and cannot accurately identify, so it has some limitations in the need of real-time monitoring scene.

Single-stage algorithms such as you only look once (YOLO) and single shot multibox detector (SSD) perform end-to-end detection directly on images, achieving real-time performance while maintaining high accuracy. For example, YOLOv3 is applied to the intelligent monitoring of illegal operations at the

construction site of construction projects to identify behaviors such as not wearing safety helmets and working clothes, with an accuracy rate of 92.17% [6]. The improved X-YOLOv11 model introduces the cross-channel partial self-attention (C2PSA) module and the complete intersection over union (CIoU) loss function, which improves the average accuracy compared with similar models, and the detection speed is as high as 145 frames per second (FPS) while maintaining a high mean average precision (mAP) [7]. The SSD model can be used for monitoring small relaxation damage in bolt connections, and calculate the relaxation angle of bolt by identifying the bolt and the number mark on it.

2.2 Method Based on Pose Estimation

Human pose estimation is the key point to locate human joints from images or videos, which is used to understand the action intention of construction personnel [8]. The target detection method is mainly used to detect static protective equipment, and the pose estimation can identify dynamic unsafe behaviors such as climbing, falling and illegal operation by capturing the spatial position and motion trajectory of human bone nodes.

According to different output dimensions, pose estimation can be divided into 2D pose estimation and 3D pose estimation. 2D method is used to predict the coordinates of joints on the image plane. It has the advantages of small amount of calculation and high real-time performance, and is suitable for edge deployment. 3D method can analyze the depth information of joints more deeply and describe the motion posture more accurately, but the computational complexity is high. Table 1 compares the performance characteristics of typical 2D and 3D pose estimation methods from the four dimensions of core ideas, advantages, limitations and application scenarios.

Table 1: Comparison of Typical 2D and 3D Pose Estimation Methods

Evaluation dimension	JC-STNet (2D real time pose estimation)	PAG-GNNA (3D pose estimation+action recognition)
Core idea	Joint level independent modeling combined with kinematics/dynamics prior	Multimodal multi flow architecture, combined with graph neural network and attention mechanism
Advantages	Lightweight, strong real-time performance, physical rationality and good robustness	High accuracy, strong feature discrimination and good generalization performance
Limitations	Weak global context modeling capability	Complex model and high computational cost
Application scenarios	Edge computing, virtual avatar, real-time 2D pose estimation	High precision motion recognition and 3D pose analysis

It can be seen from table 1 that in 2D pose estimation, joint-centric spatio-temporal network (JC-STNet) is a lightweight spatio-temporal inference framework centered on joints, which obtains the temporal motion dynamics through each joint branch, and ensures the physical rationality of pose prediction by combining kinematics in physics [9]. This method not only maintains the real-time performance, but also significantly improves the robustness in the real scene. In terms of 3D pose estimation and action recognition, PAG-GNNA method is a multimodal design framework, which combines pose estimation with target detection, models the motion and pose data in the skeleton flow through graph neural network, and introduces a learnable graph representation to enhance the robustness and discrimination of features, which greatly improves the accuracy of the framework [10].

2.3 Method Based on Time Series Modeling

Relying solely on target detection and pose estimation, the understanding of the dynamic changes of continuous actions such as climbing, crossing, and falling from height is still not deep enough. The temporal modeling method realizes the identification of dynamic unsafe behaviors by analyzing the motion patterns and temporal evolution rules between consecutive multiple frames.

SlowFast network is one of the representative temporal modeling frameworks in the field of action recognition. The framework processes video simultaneously with different frame rates through two paths. The slow path acquires spatial semantics and static scene information at a low frame rate, while the fast path recognizes dynamic features of fine motion and rapid change at a high frame rate. For example, SlowFast-Light Net introduces lightweight feature extraction network and model convergence acceleration strategy, which significantly reduces the amount of parameters and achieves 93.80% and 80.00% recognition accuracy

on the University of Central Florida 101 (UCF101) and Human Motion Database 51 (HMDB51) data sets respectively, which can provide reference for real-time behavior early warning on site [11]. Multimodal SlowFast combines slow and fast paths with 2D pose data to build a three-way architecture, which greatly improves the recognition accuracy [12].

Sequential action detection is an extended task of temporal modeling, which is responsible for accurately locating the start and end times of actions from long videos and identifying their categories. The construction monitoring video is a typical uncut long video, which needs to accurately locate the specific period of unsafe behavior in a large number of pictures, which puts forward higher requirements for the accuracy and efficiency of action detection. From the perspective of monitoring methods, the existing sequential action detection methods can be divided into fully supervised methods and weakly supervised methods. The fully supervised method can accurately demarcate the time sequence boundary, with high detection accuracy but high labeling cost. The weak supervision method only needs category labels at the video level, which has low annotation cost but relatively limited detection accuracy [13]. From the perspective of design methods, there are three types: anchor box based, boundary prediction based and query based. The anchor box based method uses pre-defined time-domain window for dense sampling, which has high recall rate but high computational cost. The method based on boundary prediction directly regresses the start and end time of the action, which is more efficient, but the processing ability of fuzzy boundary is insufficient. The query based method uses the attention mechanism to adaptively locate the action interval, which has strong flexibility but high model complexity [14].

2.4 Lightweight and Edge Deployment Methods

Construction safety monitoring focuses on real-time, while high-precision models usually have large parameters and high computational overhead, which are difficult to operate efficiently under resource constraints. Therefore, lightweight network design and edge deployment optimization become key issues.

In terms of lightweight algorithm design, researchers have proposed a variety of optimization strategies for different scenarios. In view of the problems such as the sharp change of light in the confined space of the power station and the difficulty of small target detection, GSA-YOLO embedded the GCA-C2f module integrating GhostConv and convolutional block attention module (CBAM) attention mechanism in the backbone network to reduce the amount of parameters and calculation cost and enhance the ability of feature extraction. The improved efficient channel attention (ECA) attention mechanism is introduced in the neck to suppress the characteristics of unrelated channels, and P2 detection branch is added in the head to improve the problem of small target missing detection in low light [15]. In the complex environment, the safety helmet detection accuracy is limited and the false detection and leakage detection are frequent. FGP-YOLOv8 uses FasterNet to replace YOLOv8n backbone network, introduces global spatio-temporal attention (GSTA) module to enhance feature fusion and reduce computational complexity, and uses ParNet-C2f module to replace the original C2F module to improve the multi-scale target extraction ability [16].

In terms of edge deployment framework, MMDet-Edge proposed an edge optimized unified detection framework for security critical scenarios, which achieved a balance between accuracy and efficiency through three collaborative innovations, and improved the average accuracy of small targets. In the field deployment of the three active construction sites, the framework realizes the real-time detection of five key safety targets, including personnel, helmets, flames, smoke and vests, under extreme conditions, so as to reduce safety accidents [17].

3. Common Data Sets and Evaluation Indicators

3.1 Common Datasets

3.1.1 Common Public Datasets

Public data sets are the basis for training and evaluating deep learning models. In the field of visual recognition of unsafe behaviors of construction personnel, the existing data sets can be divided into target detection data sets and action recognition data sets according to the task type. The target detection data set is

mainly used to detect the wearing of protective equipment, while the action recognition data set is used to identify and locate the dynamic unsafe behavior.

In terms of target detection tasks, construction personal protective equipment (PPE) data set is one of the most commonly used personal protective equipment detection data sets. It is published by Ultralytics and contains 11 categories, which are divided into positive (PPE worn) and negative (missing PPE). This two-way structure enables the model to simultaneously detect correctly worn equipment and identify safety violations. This data set is derived from the real construction environment and covers the marking of protective equipment such as helmets, vests, gloves, boots, goggles and so on. It is widely used in the training and evaluation of YOLO series models [18].

In terms of action recognition and behavior understanding tasks, the construction action recognition dataset contains 1595 video clips, which are specially used for video recognition research of unsafe behaviors in construction projects. It not only promotes the research of static scene understanding, but also includes dynamic activity analysis and time risk modeling in foundation pit engineering, and solves the problem of scarcity of professional data sets that can capture the changing characteristics of construction sites [19]. In addition, the split unsafe behavior recognition (UBR) data set is a benchmark data set specially built for the identification of continuous unsafe behaviors in construction scenes, which supports the model to continuously learn new unsafe behaviors without forgetting existing categories [20]. Figure 1 shows a representative sample of some unsafe behavior categories in the dataset, covering typical violation scenes such as climbing, crossing, and not wearing safety helmets. Each picture is annotated according to the site safety standards.

Figure 1: Representative samples of different unsafe behavior categories from split UBR dataset

3.1.2 Data Set Status and Deficiencies

Although the above data sets provide important support for the identification of unsafe construction behaviors, there are still the following deficiencies.

Despite the valuable contributions of these datasets, several limitations remain. First, scene coverage is limited, as most datasets focus on PPE detection and lack sufficient samples of dynamic unsafe behaviors such as climbing, crossing, and working at heights. Second, cross-regional applicability is insufficient, as differences in shooting angles, lighting conditions, and scene complexity across datasets can cause significant performance degradation when a model trained on one dataset is deployed at another site. Third, annotation costs remain high, as behavior recognition and pose estimation require frame-by-frame labeling of keypoints or action intervals, restricting the construction of large-scale, finely annotated datasets.

3.2 Evaluation Index System

3.2.1 Evaluation Index Based on Target Detection Method

Precision and recall are the two most basic evaluation indicators. The accuracy rate measures the proportion of samples that are predicted to be positive, and reflects the accuracy of the model. Recall rate measures the proportion of samples that are actually positive and are correctly detected, reflecting the completeness of the model. The calculation formula is:

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN} \tag{3-1}$$

Where TP is the number of targets correctly detected, FP is the number of false detections (predicting negative classes as positive ones), and FN is the number of missed detections (predicting positive classes as negative ones). The accuracy rate is applicable to scenarios sensitive to false detection, and the recall rate is applicable to scenarios sensitive to missed detection.

Average precision (AP) and mAP are the standard evaluation indicators in the field of target detection. AP is the area under the accuracy recall curve, reflecting the performance of the model in a certain category. mAP is the average value of all categories of AP, reflecting the overall performance of the model. The calculation formula is:

$$\text{AP} = \int_0^1 P(R)\, dR, \text{mAP} = \frac{1}{N}\sum_{i=1}^{N} \text{AP}_i \tag{3-2}$$

Where n is the number of categories. In the actual evaluation, mAP@0.5 Represents the mAP calculated when the intersection over union (IoU) threshold is 0.5, mAP@0.5:0.95 represents the average value of mAP in the range of IoU threshold from 0.5 to 0.95 (step size 0.05), and the latter has more stringent requirements on positioning accuracy. In construction safety monitoring, GSA-YOLO uses mAP@0.5 as the main evaluation index, achieving 91.2% on a self-built dataset; MMDet-Edge also uses mAP@0.5 to evaluate edge detector performance.

3.2.2 Evaluation Index Based on Pose Estimation Method

The percentage of correct key points (PCK) is a common indicator of 2D pose estimation, which is defined as the proportion of samples whose distance between the predicted key points and the real key points is less than the set threshold. The threshold is usually normalized by the ratio of the diagonal length of the head bounding box (such as 0.1, 0.2). PCK@0.1 indicates the proportion of joint points whose error is within 0.1 times the length of the head.

The mean joint position error (MPJPE) is the basic index of 3D pose estimation, which directly calculates the Euclidean distance between the predicted joint points and the real value. Its variant PA-MPJPE eliminates the influence of global rigid body transformation through Procrustes analysis, and pays more attention to the accuracy of posture itself. In the research of 3D human pose estimation, MPJPE often reports errors in millimeters.

3.2.3 Evaluation Index Based on Time Series Modeling Method

tIoU is the core index in sequential action detection, which measures the degree of overlap between the predicted action interval and the real labeled interval on the time axis. Similar to spatial IoU, tIoU is calculated by dividing the intersection length of the predicted interval and the real interval by the union length.

mAP is the most commonly used evaluation index in sequential motion detection. For a given tIoU threshold, the AP of each action category is calculated first, and then the mAP is obtained by averaging all categories. The final evaluation index is usually the average value of mAP in the range of tIoU threshold from 0.5 to 0.95 (step size 0.05).

3.2.4 Evaluation Index Based on Lightweight and Edge Deployment Method

Reasoning speed (FPS), which measures the number of image frames processed by the model per second, is the core index to evaluate the real-time performance. The higher the FPS, the faster the model processing

speed. When deploying edge devices, it is also necessary to pay attention to the latency of end-to-end reasoning, including preprocessing, inference, and post-processing.

Parameters and floating-point operations (FLOPs) are the key indicators to measure the lightweight degree of the model. Parameters reflect the storage requirements of the model, and FLOPs reflect the computational complexity of the model. The smaller the two, the more suitable the model is for deployment on resource constrained edge devices. In the evaluation of lightweight algorithm, the amount of parameters, calculation and detection accuracy are usually reported synchronously to comprehensively measure the performance tradeoff ability of the model under resource constraints.

4. Conclusions and Outlook

4.1 Conclusions

In terms of technical route, the behavior recognition method based on deep learning has developed from the previous static detection to the real-time dynamic recognition detection. YOLO series, Faster R-CNN and other target detection methods are mainly used to detect static protective articles such as helmets and reflective clothing, while the single-stage detector with real-time advantages is widely used in large-scale monitoring scenes. JC-STNet, PAG-GNNA and other pose estimation methods realize the understanding of the meaning of climbing, falling and other movements by grasping the key points of human bones, and progress from "seeing" to "understanding". Multimodal SlowFast and temporal modeling methods such as temporal motion detection framework analyze motion patterns between consecutive frames, accurately locate and identify dynamic unsafe behaviors. GSA-YOLO, FGP-YOLOv8, MMDet-Edge and other lightweight and edge deployment methods pursue the balance of accuracy and efficiency, and promote the technology from the laboratory to the site.

In terms of data sets and evaluation indicators, the existing public data sets provide basic support for model training, but the scale is generally small, the scene coverage is limited, and there is still a significant gap relative to the complexity of the real site environment. mAP, FPS, and parameter quantity have formed a relatively complete evaluation system, which provides a quantitative basis for the fair comparison of different methods.

Although the current method has achieved remarkable results in specific scenarios, for example, MMDet-Edge system has achieved a significant reduction in safety accidents in field applications. However, the existing research still faces many challenges, including the difficulty of small target detection, the vulnerability to interference in complex environments, the lack of robustness, and the difficulty in balancing real-time performance and accuracy.

4.2 Outlook

In response to the above challenges, future research can explore the following directions.

In terms of multimodal fusion, single visual recognition is easy to be affected by harsh lighting and occlusion conditions. In the future, multi-source sensor data such as infrared thermal imaging, depth camera and so on should be fused to realize multi-mode collaborative recognition of vision and sensing, and improve the recognition reliability in complex environments.

In terms of large model empowerment and small sample learning, the current method relies on a large number of fine annotation data, which is expensive and difficult to cover various types of work and scenarios. In the future, large visual language models (such as contrastive language-image pre-training (CLIP) and generative pre-trained transformer 4 with Vision (GPT-4V)) can be introduced to realize behavior recognition and semantic interpretation with zero or few samples, so that the model can understand the complex safety rules described in natural language, such as "workers entering hazardous areas without safety helmets", and quickly adapt to new scenarios in combination with prompt learning strategies, significantly reducing the cost of model migration.

In terms of lightweight and hardware co design, the edge equipment of the construction site needs the model to pursue speed while maintaining accuracy. Future work should explore the collaborative optimization strategy of automated neural architecture search and hardware perception, develop dynamic reasoning

mechanism, and automatically adjust the allocation of computing resources according to the complexity of the scene, so as to achieve the dynamic balance between accuracy and speed. The hardware algorithm collaborative design path of MMDet-Edge and the lightweight practice of GSA-YOLO provide examples for this.

To sum up, the deep learning driven identification of unsafe construction behaviors is changing from traditional static identification to accurate dynamic real-time monitoring. Through the in-depth integration and continuous optimization of multi-technology paths, it is expected to realize the comprehensive intelligent upgrading of construction site safety management.

References

[1] National Bureau of Statistics. (2025) National Bureau of Statistics: A total of 19,626 people died in various production safety accidents in 2024. Safety Management Network. Retrieved from https://www.safehoo.com/Case/Stat/202503/5764614.shtml

[2] Tong, R. P. (2019) Review and commentary on behavioral safety research progress. Safety, 40(7), 1-14+88.

[3] Wu, C., & Wang, B. (2018) Research on meta-model of behavioral safety management. Journal of Safety Science and Technology, 14(2), 5-11.

[4] Fang, W. L., & Ding, L. Y. (2022) Research on intelligent recognition and correction of workers' unsafe behaviors. Journal of Huazhong University of Science and Technology (Natural Science Edition), 50(8), 131-135.

[5] Zhang, Y. (2020) Research on key technologies of intelligent safety monitoring for building structure construction. Dalian University of Technology.

[6] Liu, X. J. (2023) Intelligent monitoring method for illegal operations at construction sites based on machine vision. China Construction Informatization, (10), 82-86.

[7] Wang, Z. F., Yang, T., Min, H., et al. (2025) Research on mine fire image recognition method based on deep learning model. Journal of Safety Science and Technology, 21(8), 146-154.

[8] Kappan, M. M., Sandoval, B. E., Meijering, E., et al. (2025) A survey on deep learning for 2D and 3D human pose estimation. Artificial Intelligence Review, 59(1), 32-32.

[9] Cai, X., Qu, K., Liu, C., et al. (2026) JC-STNet: a physics-inspired joint-centric spatiotemporal network for real-time 2D pose estimation. The Visual Computer, 42(9), 365-365.

[10] Baghel, A., Kushwaha, S. K. A. (2026) PAG-GNNA: Advanced Graph Neural Network Approach for 3D Human Pose Estimation and Action Recognition. International Journal of Pattern Recognition and Artificial Intelligence.

[11] Xu, Y., Lu, Y. (2025) Research on Action Recognition Algorithm Based on SlowFast Network: Regular Papers. Journal of Advanced Computational Intelligence and Intelligent Informatics, 29(5), 1056-1061.

[12] Lasri, K., Fazazy, E.K., Mahraz, M. A., et al. (2026) Video action recognition based on multimodal SlowFast architecture. Results in Engineering, 31111265-111265.

[13] Junyi, S., Ma, L., Jikai, Z. (2022) Temporal Action Detection Methods Based on Deep Learning. International Journal of Pattern Recognition and Artificial Intelligence, 36(3).

[14] Kai, H., Chaowen, S., Tianyan, W., et al. (2024) Overview of temporal action detection based on deep learning. Artificial Intelligence Review, 57(2).

[15] Wang, H., Zhou, Q., Hao, Z., et al. (2026) Lightweight Safety Helmet Wearing Detection Algorithm Based on GSA-YOLO. Sensors, 26(7), 2110-2110.

[16] Zhang, L., Ma, H., Huang, J., et al. (2025) An Improved Lightweight Safety Helmet Detection Algorithm for YOLOv8. Computers, Materials & Continua, 83(2), 2245-2265.

[17] Zhu, T., Liu, H., Duan, H., et al. (2026) MMDet-Edge: A Multi-Scale and Multi-Object Detection Framework for Safety-Critical Edge Deployment. Sensors, 26(4), 1151-1151.

[18] Dalvi, M., Singh, N., Bhingarde, S., et al. (2025) Construction-PPE: Personal Protective Equipment Detection Dataset.

[19] Zhao, C., Ye, W.W., Xing, J.Q., et al. (2025) A Comprehensive Image and Video Dataset for Computer Vision-Based Safety Monitoring in Foundation Pit Construction. Mendeley Data.

[20] Wang, T., et al. (2025) Rehearsal-Free Continual Learning for Emerging Unsafe Behavior Recognition in Construction Industry. Sensors.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access