International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 8159-8169

https://doi.org/10.70517/ijhsa463697

A Study on the Application of Automation Technology to Improve the Efficiency of Civil Engineering Project Management

Yongming Zhou[1,*]
[1] Guangxi Transportation Industry Co., LTD., Nanning, Guangxi, 530000, China
Corresponding authors: (e-mail: 17776263546@163.com).

Abstract The scale of modern civil engineering projects has been expanding, and traditional management methods have been difficult to meet the dual requirements of efficiency and safety. This study explores the application of automation technology in civil engineering project management and constructs an unsafe behavior identification model based on deep learning. Through the integration of on-site monitoring system and schedule control technology, it realizes accurate quality management and safety risk warning of engineering projects. The study designs an unsafe behavior recognition model based on CNN-LSTM, which combines the Inception-v3 framework to extract spatial features, and captures temporal dynamic features through a two-layer LSTM network to realize the intelligent recognition of workers' unsafe behaviors. The experiments use UCF-101 public dataset and self-constructed construction site dataset for model training and validation, and the results show that the model has a recognition accuracy of 94.52% on the UCF-101 dataset, with a computational complexity of 8.28G, and a parameter count of only 6.16M, which is a 5%-11% increase in the average accuracy value compared with that of the traditional methods. In actual engineering applications, the system collected more than 1,200 pieces of information on personnel's “three violations” in one year, effectively reducing the incidence of unsafe behaviors on site. The study proves that automation technology combined with deep learning model can effectively improve the safety management efficiency of civil engineering projects and provide technical support for intelligent supervision of engineering.

Index Terms automation technology, civil engineering, project management, CNN-LSTM model, unsafe behavior recognition, deep learning

I. Introduction

Civil engineering construction is developed along with the development of human society, which directly responds to the development of the current society's economy, culture and science and technology, etc. [1]. In recent years, with the booming development of the construction market, the society has also put forward higher requirements for civil engineering construction. More and more new technologies, new materials, new techniques, new institutions are integrated into the modern civil engineering industry, the shape of the building, institutions, construction techniques, etc. are changing day by day, and the scale of the building is getting bigger and bigger, the structure is getting more and more complex, and the form is getting more and more diversified [2]-[5].

With the current development trend of civil engineering, the application of project management in civil engineering building construction is also more and more extensive and comprehensive [6], [7]. Through strict project management work, it can macroscopically coordinate the resource allocation of each project and improve the efficiency of resource use [8], [9]. It can be seen that project management work is an effective guarantee for civil engineering construction to achieve economic and social benefits [10].

However, civil engineering construction project management, as a complex production activity, involves numerous human, material and technical issues, in order to be able to ensure the smooth progress of the whole project, it is necessary to carry out scientific management of geotechnical engineering in order to improve the efficiency of civil engineering project management [11]-[13]. Based on this, the introduction of automated computing in the process of civil engineering project management can not only improve the utilization efficiency of resources, but also effectively control the environmental pollution problems at the construction site.

Currently, the civil engineering field is experiencing profound digitalization changes, and the complexity and scale of engineering projects continue to grow, which puts forward higher requirements for management efficiency and safety and security. Traditional project management methods rely on manual empirical judgment and paper document records, which have obvious limitations in information transfer, resource deployment and risk prediction.

Information silos, communication lags and blind spots in safety supervision in project site management have become key factors restricting project quality and progress. Data show that the rate of schedule delay due to mismanagement in the building construction industry is as high as 40%, and safety accidents also remain high. In this context, exploring the innovative application of automation technology in civil engineering project management has become an important direction to solve the industry's pain points.

Automation technology creates a new paradigm for engineering management through digital sensing, intelligent analysis and automatic control. IoT sensors can monitor the state of engineering structures in real time, BIM technology realizes the integration and sharing of engineering whole life cycle information, and artificial intelligence provides data support for decision-making. The integration and application of these technologies significantly improves project transparency and traceability, enabling managers to identify and solve various problems in the construction process in a timely manner. Especially in the field of safety management, computer vision-based behavior recognition technology can automatically detect and warn of unsafe behaviors and prevent safety accidents from the source.

Intelligent safety supervision is gradually becoming a core component of engineering project management. By analyzing accident cases, it can be seen that workers' irregular operation and violation of safety regulations are the main reasons for accidents. Traditional safety inspection methods have limited coverage and time lag, making it difficult to realize comprehensive monitoring. Deep learning technology can automatically identify dangerous behavior patterns from massive video data, providing technical support for safety management. However, the existing research mainly focuses on the behavior recognition in general scenes, and the research on the recognition of unsafe behaviors in the special environment of construction sites is relatively insufficient.

Based on the needs of civil engineering project management, this study systematically analyzes the effectiveness of the application of automation technology in site monitoring, quality management and safety production, and focuses on constructing an unsafe behavior recognition model based on CNN-LSTM. The model integrates the advantages of deep convolutional neural network and long and short-term memory network, which can extract spatial features and time series features of behaviors at the same time, realizing accurate identification of unsafe behaviors such as workers not wearing helmets and overtopping guardrails. The study constructs a professional dataset through a variety of data collection methods, adopts a transfer learning strategy for model training, and conducts a comprehensive assessment and comparative analysis of model performance. Finally, the study selects actual engineering projects for application verification to assess the actual effect of the model in improving the efficiency of safety management, providing theoretical basis and technical support for intelligent management of civil engineering projects.

II. Application of automation technology in civil engineering project management

With the accelerating process of urbanization and infrastructure construction, the traditional construction management mode has been difficult to meet the double requirements of efficiency and safety. In this context, automation technology has gradually become an important engine to promote the digital transformation of civil engineering management, which is effective in improving the efficiency of on-site decision-making and resource utilization, and can help all parties to share information and collaborative management on the basis of the overall improvement of project quality and efficiency.

II. A. Site monitoring and progress control

Automation technology has significant advantages in on-site monitoring and progress control. Through the deployment of various sensors, cameras and intelligent terminal devices, the operation area is digitally upgraded, making the construction project more visible and controllable, and laying a solid foundation for the safe and steady advancement of the project progress. Firstly, IoT devices can transmit real-time data information such as concrete pouring temperature, pit deformation, and steel reinforcement arrangement back to the cloud management platform, enabling managers to synchronize the view of construction dynamics in the office or on the mobile and provide timely warnings of abnormal situations. Secondly, by integrating GIS maps and BIM, all parties involved in the project can locate key nodes in a three-dimensional visualization environment, track the progress of each construction stage and conduct data comparison and analysis, thus effectively reducing information lag and communication barriers. In addition, the use of intelligent image recognition and video monitoring technology automatically detects whether workers are wearing safety helmets in a standardized manner and whether they are working in accordance with standard procedures, providing multiple guarantees for project safety management and quality control. Relying on this integrated platform, project managers can not only obtain complete construction logs, but also make reasonable forecasts of the construction period and timely adjustments to resource allocation with the help of big data analysis.

II. B.Engineering Quality Management and Intelligent Inspection

Intelligent detection means can realize the process of dynamic tracking and instant feedback. First of all, with the help of wireless sensors, the concrete strength, humidity, vibration situation for real-time recording, to ensure that in the pouring and curing stage can be accurate data support, which is more critical for high-rise buildings or special process requirements of the construction site. At the same time, the installation of stress-strain monitoring devices in the formwork and bracket positions can timely detect structural overload or deformation potential hazards, so that managers can immediately take corrective or reinforcing measures. Secondly, based on the visual comparison and modeling function of BIM, technicians can quickly locate construction errors or material defects, scan and analyze the deviated areas and propose targeted repair solutions. In addition to the intelligent detection of hardware, software algorithms also play an indispensable role. Utilizing data mining and AI image recognition technology, the system can automatically compare the design drawings with the actual construction conditions, issue red flags for deviations from the standard and generate quality alerts. Comprehensively speaking, automation technology provides a higher level of quality assurance for the construction process through accurate quality monitoring and automated detection means, and also lays a credible digital foundation for subsequent acceptance work.

II. C.Safe production and early warning of risks

First, RFID or face recognition equipment is equipped at construction sites, which can record the identity and distribution of people entering and leaving the site in real time. When overcrowding or unauthorized personnel intrude in certain operation areas, the system will automatically prompt management personnel to intervene. Secondly, for potentially high-risk segments such as overhead work and large equipment operation, intelligent video surveillance combined with AI algorithms can be deployed to identify and analyze workers' behavioral postures, and once a failure to wear a seatbelt or a violation of standing position is detected, an alert can be sent in a timely manner. In addition, in terms of construction machinery, load sensing monitoring of tower cranes, excavators and other important equipment can predict the degree of equipment fatigue and the risk of potential failures, and send early warning messages before the threshold is critical. The environmental monitoring module of the smart construction site also provides instant monitoring of toxic gas levels, dust concentrations and meteorological changes, providing a scientific basis for special construction processes or health risk control. By combining real-time data with early warning models, managers are able to effectively deal with accidents in their infancy, thereby reducing workplace injuries and property damage. As the technology continues to mature, the intelligent prevention and control capabilities of automation technology in terms of production safety and risk early warning will continue to increase, injecting more vitality into the all-round safety and security of construction sites.

III. CNN-LSTM based unsafe behavior recognition model

In this paper, we construct an unsafe behavior recognition model based on CNN-LSTM to effectively identify and warn the unsafe behaviors of the personnel in the construction area of the project, so as to reduce the occurrence of safety accidents and improve the construction efficiency and construction quality of the construction site.

III. A. Long- and short-term memory networks

Recurrent Neural Networks (RNN) are an extension of traditional feed-forward neural networks with the ability to manage variable long sequences of inputs. Long Short-Term Memory Networks (LSTM) is an extension of RNN which is capable of solving the gradient vanishing problem in a very clean manner.The LSTM model essentially extends the memory of the RNNs so that they are able to maintain long term dependencies with the learning inputs.This extension of the memory has the ability to memorize the information for long periods of time and hence is able to read, write, and erase the information from the memory.The LSTM model is an extension of the RNN which has the ability to manage long sequences of inputs.

In general, an LSTM model consists of three gates: an oblivion gate, an input gate, and an output gate. The LSTM memory is known as a gating unit, and the term “gate” is inspired by the ability to decide whether to save or ignore memorized information. The LSTM model captures important features from the inputs and saves them over a long period of time, and the decision to delete or retain information is based on the weight values assigned to the information during the training process. .

The forgetting gate decides whether to keep or delete existing information, the input gate specifies the extent to which new information will be added to memory, and the output gate controls whether existing values in the cell contribute to the output. Oblivion gates typically use a sigmoid function to decide what information needs to be removed from LSTM memory. This decision is essentially made based on the values of h_{t-1} and x_t. The output of the forgetting gate is f_t, a value between 0 and 1, where 0 means completely removing the learned value and 1 means keeping the entire value. This output is calculated as:

$$f_t = \sigma\left(W_f[x_t, h_{t-1}] + b_f\right) \tag{1}$$

where x_t is the input vector at the current moment, h_{t-1} denotes the output vector at the previous moment, b_f is the bias, W_f is the weight matrix, and σ is the sigmoid, the activation function of the forgetting gate. the output of the forgetting gate f_t is designed to be multiplied by C_{t-1}, and C_{t-1} is the memory of the output of the previous moment, which means that the content of C_{t-1} is to be completely forgotten if the value of f_t is zero.

The input gate controls the content of the next moment added to the LSTM memory. The gate consists of two layers: a sigmoid layer and a tanh layer. the sigmoid layer activates the information value that reaches the threshold, and the tanh layer creates a vector of new candidate values and adds them to the LSTM memory. The outputs of these two layers are computed:

$$\begin{cases} i_t = \sigma\left(W_i[h_{t-1}, x_t] + b_i\right) \\ \tilde{C}_t = \tanh\left(W_C[h_{t-1}, x_t] + b_C\right) \end{cases} \tag{2}$$

$\tilde{C}_t$ can be viewed as the current input to the network update, and the tanh activation function normalizes the new information. And i_t is similar in structure to the forgetting gate and is used to decide how much of the new content needs to be retained. The combination of the input gate with the forgetting gate output gate updates the current cell state, i.e., generates C_t:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{3}$$

The output gate first uses a sigmoid layer to decide which part of the LSTM memory to use for output, and then performs a nonlinear hyperbolic function to map the cell state between -1 and 1. Finally, the result is output through the sigmoid layer:

$$\begin{cases} o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \\ h_t = o_t * \tanh(C_t) \end{cases} \tag{4}$$

LSTM is able to model long sequences of inputs and outputs by controlling the updating and forgetting of cell states.

III. B. CNN-LSTM modeling

The CNN-LSTM based unsafe behavior recognition model used in this work is shown in Fig. 1. The whole model is divided into a two-layer structure: firstly, a convolutional neural network is used to compute the preprocessed action video and extract its feature representation. Then, the time-varying feature sequences are extracted using the LSTM model and inserted into the Softmax classifier in the last layer of the CNN architecture to complete the behavior recognition.

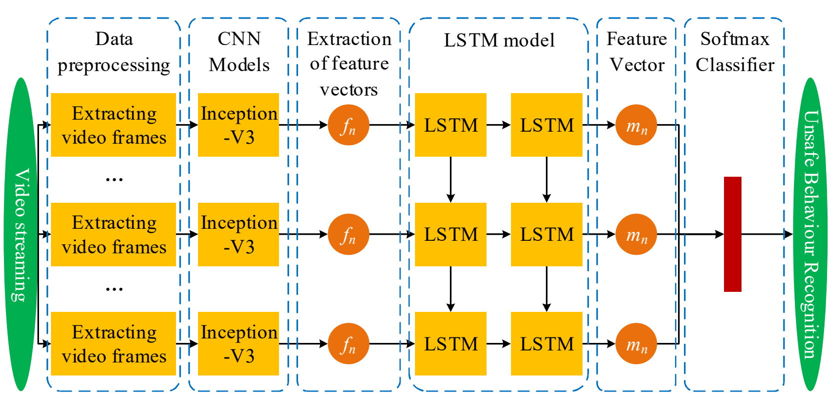

Figure 1: Unsafe behavior recognition model based on CNN-LSTM

The Softmax function used in the classification process is represented in the form of a probability function as shown in equation (5):

$$P\left(y^{(i)}=n \mid x^{(i)};W\right)=\begin{bmatrix} y^{(1)}=1 \mid x^{(1)};W \\ y^{(2)}=n \mid x^{(2)};W \\ \vdots \\ y^{(i)}=n \mid x^{(i)};W \end{bmatrix} = \frac{1}{\sum_{j=1}^{n} e^{W^T x^{(i)}}}\begin{bmatrix} e^{W_1^T x^{(i)}} \\ e^{W_2^T x^{(i)}} \\ \vdots \\ e^{W_n^T x^{(i)}} \end{bmatrix} \tag{5}$$

where P is the i rd training sample out of m, the j th class out of n with weight W, and $W_j^T x^{(i)}$ denotes the input to the Softmax layer.

The implementation process of this model is as follows:

(1) Segment the action video and randomly select 25 video frame segments as representatives of the time nodes. The video frames are fed into the CNN model, the development framework of the CNN model in this paper uses Inception-v3, after which the convolutional layer calculates the output of each neuron that is connected to a localized region for each input video. Finally the dot product between their weights and the small regions they are connected to is calculated so that the CNN output from the fully connected layer will be a 2048 dimensional feature vector that will be used as a spatial feature of the action video.

(2) Repeat the operation of the first step to obtain the spatial features of each video frame, i.e., 25 feature vectors can be obtained for one video. These feature vectors are fed into the LSTM model for temporal feature training. Note that the generation of temporal features starts from the second LSTM model generated.

(3) Input all the temporal features into the Softmax classifier and average them over the time period to obtain a probabilistic representation.

Layer 1, deep CNN model: the CNN has a multi-layer architecture, which is capable of automatic feature extraction, and can also utilize a classifier to map the extracted feature vectors to the final prediction. For each layer in the CNN model, a convolution operation and a function operation are performed on the output of the previous layer in the forward propagation stage, as shown in equation (6):

$$h_{ij}^{k}=f\left(\left(W^{k}x\right)_{ij}+b_{k}\right) \tag{6}$$

where f is the activation function, b_k is the offset of this feature mapping, and W^k is the value connected to the k th feature mapping.

In the Inception-v3 framework, the module consists of a composite layer containing five uniformly shaped filters including 1×1, 3×3, and 5×5 sized convolutional kernels and 3×3 sized outputs of average pooling operations. In this paper, the Inception-v3 framework with pre-training parameters based on ImageNet is directly applied to the action video frames.ImageNet is a large visual database for visual recognition software. The output size of the last pooling layer is chosen to be 1 × 1 × 2048 as the spatial features of the output, i.e., the 2048-dimensional features generated by the last pooling layer in the forward pass of the pretrained model are used. As a result, the model does not retrain and repeat the initial network through backpropagation, and no retraining of the model is required to significantly reduce training time.

In the unsafe behavior recognition model, the LSTM model is divided into a two-layer network architecture to handle the temporal dynamics of the video features generated by the last pooling layer of the Inception-v3 network. $\{f_1, f_2, \cdots, f_n\}$ are the n features computed by the Inception-v3 network from each video n frame. Thus, for each input sequence $\{f_1, f_2, \cdots, f_n\}$, the storage units in the two LSTM layers will generate a sequence of representations $\{m_1, m_2, \cdots, m_n\}$. This sequence is then averaged over the entire time period to obtain a temporal feature vector F, denoted as:

$$F = \frac{m_1 + m_2 + \cdots + m_n}{n} \tag{7}$$

After that this feature vector F is fed into the Softmax layer so that the unsafe behavior in each input video can be identified.

IV. Data set construction and model analysis

IV. A. Data acquisition and processing

IV. A. 1) Basic data collection

Data on unsafe behaviors at construction sites are usually sensitive, and it is difficult for researchers to fully obtain data on construction violations at construction sites within a limited period of time. In order to fully obtain the data related to unsafe behaviors in this study, a combination of experimental simulation photography, internet search, construction site surveillance video extraction, and construction site cell phone photography was used to collect sample data.

When simulating the shooting at the experimental site, by building a simple construction site construction scene, volunteers were invited to shoot videos of wearing \ not wearing a helmet, climbing over the guardrail, and using a cell phone, respectively, and the key frames of the shooting video were obtained through professional video editing software, and the final export of the photos in JPG format.

Network search for unsafe behavior images, mainly using large search engines to search for relevant image data. The search terms mainly include “helmet”, “construction”, “safety accidents”, “workers”, “construction”, “call”, “ove”, “over guardrail”, “rainy season construction”, “safe and civilized construction”, “night construction”, etc. Download and save related images.

Construction site surveillance video extraction of worker construction images. Acquire some of the surveillance video data of a civil engineering project, and finally save the video frame data with higher quality by Edius editing.

Summarizing all the data collected in this study, a total of 2521 video frames and pictures are included. Considering the complex conditions of on-site construction environment, uncontrolled lighting, weather climate change, large changes in the spatial scale of the target, the target being obscured, and the target being blurred, in order to ensure that the detection model has sufficient sensitivity to the detection target, the data it collects should also basically cover the various situations mentioned above.

IV. A. 2) Data annotation

Matlab computer vision toolbox contains image and video data labeling tools Image Labeler and Video Labeler, the above two tools can be used to mark the target object region (ROI) for object detection. After importing the images into the labeling tool, the workers' behaviors of wearing/not wearing helmets, climbing over guardrails, and making phone calls are labeled manually to form the labeled dataset.

IV. A. 3) Data enhancement

In order to ensure that the image characteristics remain unchanged during the process of data enhancement, this study adopts the data enhancement method of online enhancement to process the acquired image data. The final data enhancement methods used in this paper include horizontal flipping, random adjustment of brightness and saturation, and Gaussian blur.

IV. B. Model Experiment Analysis

IV. B. 1) Data sets

The experimental data uses the UCF-101 dataset and the self-constructed construction site construction personnel dataset for unsafe behavior recognition.The UCF-101 video dataset is a widely used public dataset in the field of behavior recognition, which is mainly derived from the YouTube video platform, and includes 101 behavioral categories, including human-object interactions, human-human interactions, human actions, musical instrument movements, and sports. The UCF-101 dataset not only contains a variety of behavioral categories, but also contains a variety of complex backgrounds, with problems such as low light and camera shake, which have high requirements for model robustness. The self-constructed building construction personnel dataset is mainly derived from the data collection in the previous section.

IV. B. 2) Model training

The experiment is based on the PyTorch deep learning framework to build the model, the model is trained using the cross-entropy function as the loss function, which is used to measure the similarity between the classification results of the trained samples and the labels of the trained samples. The model training uses adaptive moment estimation (Adam) method to achieve the loss function convergence to the minimum value.Adam method is characterized by adaptive adjustment of the learning rate, suitable for large-scale data and parameter scenarios, simple to implement, efficient in computation, and less memory requirements. The experiment adopts the migration learning method, first based on the UCF101 dataset for model training, after the completion of training to fix some of the model parameters, the use of self-constructed building construction personnel behavior dataset for training, to obtain the final experimental results.

Figure 2 shows the loss function curve and the change curve of the recognition accuracy of the test set during the training of the proposed model on the UCF-101 dataset. With the increase of the number of training rounds, the average loss function and the average accuracy of each round tend to stabilize, and the final loss value decreases to about 0.08, and the recognition accuracy of the test set can reach up to 94.52%, which can be analyzed to show that the training results of the proposed model are more satisfactory and have a high recognition accuracy.

The model is trained twice on the self-constructed construction site dataset using the migration learning method, and the change of loss and accuracy in the model training is shown in Fig. 3, and the highest recognition accuracy of 94.76% is finally obtained on the test set. The model already has a high recognition accuracy by training on the UCF-101 dataset, and has a higher recognition accuracy on the self-constructed dataset due to the smaller amount of data, relatively single scene, and relatively fewer behavioral categories.

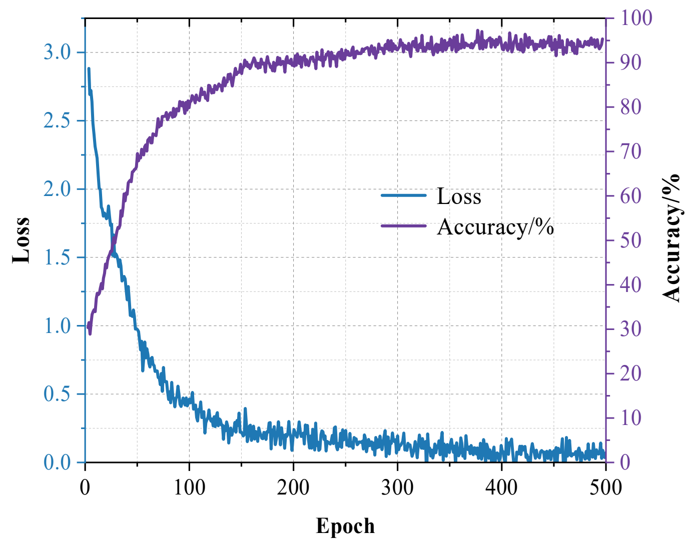

Figure 2: Changes of loss and accuracy in model training(UCF-101)

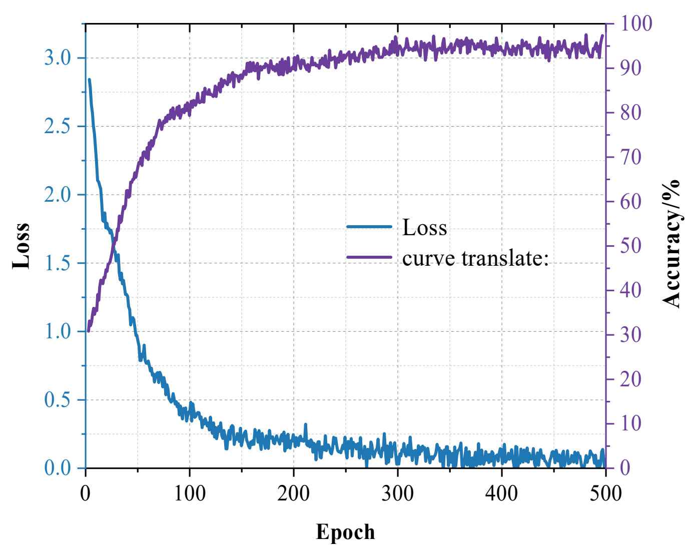

Figure 3: Changes of loss and accuracy in model training(Self-built dataset)

IV. B. 3) Evaluation of model performance

Considering that the designed CNN-LSTM insecure behavior recognition model is mainly applied to edge devices, the computational complexity of the model and the number of model parameters, in addition to the recognition accuracy, have become the main metrics for evaluating the model, which needs to be evaluated in terms of the amount of computation and size of the model. The computational complexity of the model is measured using the number of cumulative product operations, which is characterized using the number of multiplication and addition operations included in the model. The size of the model is usually characterized using the number of parameters of the model, which directly determines the size of the model file and also affects the memory footprint of the model for inference. Model accuracy is characterized by counting the proportion of correctly identified behavioral category samples to the total number of classified samples.

In order to verify the performance indicators of different methods, experiments were conducted on multiple methods. Due to the large number of samples in the UCF-101 dataset and the fact that the same behavior covers multiple scenarios, the robustness requirement of the model is relatively high. Therefore, the experiment uses this dataset to compare different algorithms. The performance comparison results of different models are shown in Figure 4. The MBH network has a large number of parameters, a low behavior recognition rate, and poor performance. The YOLOv3 model can significantly improve the recognition accuracy rate to 91.92%, but both the number of model parameters and the amount of calculation are relatively high. The recognition accuracy of the CNN-LSTM model in this paper on UCF101 is relatively high, reaching up to 93.77%. The computational complexity is 8.28G and the number of model parameters is 6.16M. Through comparison, it can be seen that both the computational complexity and the number of parameters of this model are the smallest, and it has greater advantages compared with traditional recognition methods. The comprehensive performance for identifying unsafe behaviors at civil engineering sites is better.

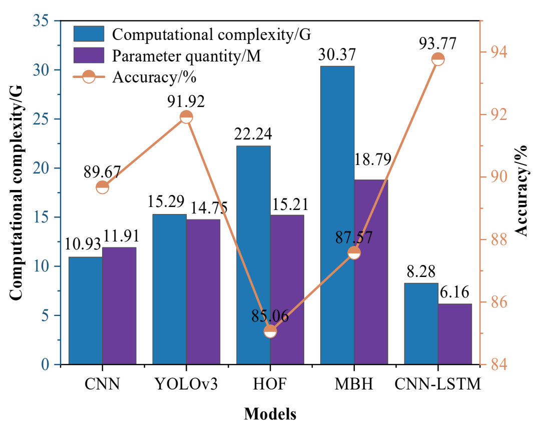

Figure 4: Performance comparison of different models

On the self-constructed construction site dataset, the average precision (AP) of different models in four scenarios, helmet, no_helmet, calling and cross_guardrail, is compared, and the comparison of AP of the five types of models is shown in Fig. 5.The average precision mean values of the CNN, YOLOv3, HOF, MBH, and CNN-LSTM models MAP are 0.78, 0.81, 0.75, 0.80, and 0.86, respectively, and the model in this paper has the highest MAP value for insecure behavior recognition, which is improved by 0.08, 0.05, 0.11, and 0.06 over the comparison methods.

In addition, the AP values of the two types of targets, helmet and No_helmet, are higher than those of calling and cross_guardrail, corresponding to 0.83 and 0.86, 0.76 and 0.74, and the AP value of calling is higher than that of cross_guardrail, which is closely related to the number of training samples for each type of target. The reason for this result is closely related to the number of training samples for each type of target. helmet and No_helmet have a sufficient number of samples, and the model fully learns the characteristics of the target corresponding to each level during the learning process, while the number of samples for calling and cross_guardrail is relatively small, and the spatial scale of the target is generally larger, so the model may not have sufficiently learned the characteristics of the samples corresponding to each level, resulting in a relatively low detection accuracy.

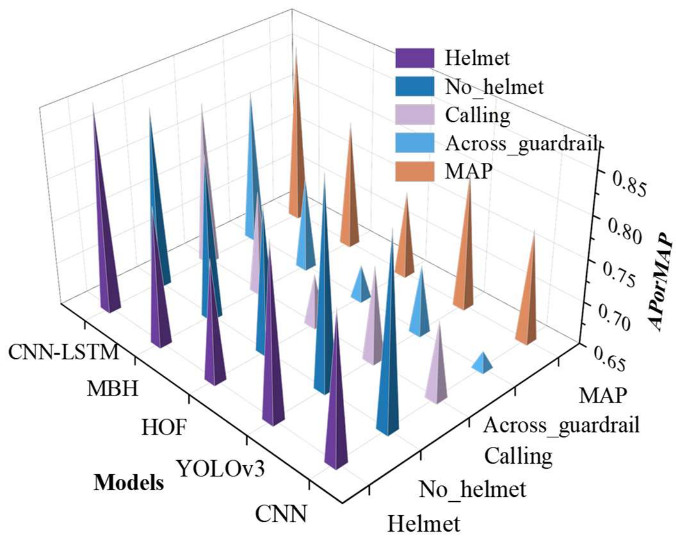

Figure 5: The comparison of AP value of the five models

In the actual project, the video frame data collected at the construction site is characterized by large variations in the lighting environment, image blurring, target occlusion, and large variations in the spatial scale of the target detection object. Therefore, this paper further classifies the test set according to different conditions and tests the overall MAP value of model target detection. The categorized test target detection results are summarized in Table 1. In terms of lighting conditions, when the lighting is balanced, the image is clear, and the features of each part of the target are rich and full, and the overall target detection MAP value of the model in the test is 0.878. When the lighting is weaker/stronger, due to insufficient or too strong lighting, the image is too dark or too bright, and the target in the image may be unclear as a result, but the target MAP value in the test also reaches 0.826, indicating that the model has basically learned the characteristics corresponding to the illumination, and has sufficient robustness to changes in illumination.

Analyzing the impact of the target occlusion situation on the detection results, when the target in the image has occlusion, the MAP is 0.783, compared with the MAP of 0.929 in the no-occlusion condition, the recognition accuracy has slipped, indicating that the model of this study has relatively limited sensitivity to the target occlusion situation. The main reason may be that the occluded targets in the collected samples are not obvious in their own characteristics, and the proportion of occluded targets to the total number of target samples is too low, which leads to the model not fully learning the various characteristics possessed by the target.

The MAP of the far-view target test in this study is 0.721, which is low in the model's far-view detection ability compared to the mid-view and near-view detection effects, which is directly related to the fact that some of the far-view targets in the training samples are too small in scale. Therefore, if this model is applied to real engineering projects, the use of long-distance and large-coverage cameras should be avoided as much as possible, and it is recommended to use medium-view and near-view cameras as a way to improve the reliability of unsafe behavior detection in civil engineering projects.

In conclusion, the CNN-LSTM-based worker unsafe behavior recognition model proposed in this study has excellent overall performance in terms of target detection accuracy and detection speed, and can be used to assist on-site monitoring in civil engineering projects.

Table 1: Classification test results summary

Test category	Particulars	Sample size (frame)	MAP
Illumination condition	Light balance	3219	0.878
	Light is weaker/stronger	1281	0.826
Occlusion	Occlusion	3128	0.783
	Unoccluded	1271	0.929
Target space scale	Far view	1317	0.721
	Middle view	926	0.908
	Near view	2224	0.945

IV. C. Model application

In order to verify the practicality of the CNN-LSTM unsafe behavior recognition model and improve the shortcomings of the model, a group carries out a trial in its building construction projects, and selects a housing construction project as a pilot unit to fully use the model in construction site safety supervision, and the system has been put into operation for one year, and more than 1,200 pieces of information on the three violations of personnel and more than 850 pieces of information on the unsafe state of objects and the unsafe state of the working environment have been collected. In the first year of operation, the system has collected more than 1,200 pieces of information on personnel's "three violations", and more than 850 pieces of information on the unsafe state of objects and the unsafe state of the working environment, which has greatly improved the efficiency of project safety supervision.

After comparing the data before and after the application of the model, it is found that the unsafe behavior of the personnel on the site of the housing construction project is gradually decreasing, and the risk factors such as the unsafe state of the objects and the environment are also gradually decreasing compared with other projects, which greatly reduces the supervisory pressure of the project safety management personnel and ensures the production safety of the project while enhancing the efficiency of the safety supervision of the project.

V. Conclusion

Deep learning-driven automation technology opens up new ways for civil engineering project management efficiency improvement. The CNN-LSTM unsafe behavior recognition model constructed in this paper shows excellent performance in complex engineering environments through the dual extraction of spatial and temporal features. The experimental data show that the model achieves 94.52% recognition accuracy on the UCF-101 dataset, and the recognition accuracy is further improved to 94.76% on the self-constructed construction site dataset. Compared with the traditional method, the model has significant advantages in terms of computational complexity and number of parameters, which are only 8.28G and 6.16M, while the average accuracy MAP value reaches 0.86, which exceeds the comparison method by 5%-11%.

The evaluation of the model's adaptability to different application conditions shows that the MAP value is 0.878 in the light-balanced environment, and the recognition accuracy of 0.826 can still be maintained in the under-/over-illumination environment; the MAP value of the target is as high as 0.929 in the case of unobstructed target, and 0.783 in the case of occluded target; and the MAP values of the near-view, medium-view and far-view conditions are 0.945, 0.908 and 0.721, respectively. this indicates that the model has strong robustness to light changes, but needs to be optimized in long-distance target recognition and occlusion situation processing.

The verification of engineering practice shows that the model has collected more than 1,200 pieces of information on "three violations" and more than 850 pieces of information on unsafe state of objects and unsafe state of the environment in one year, which effectively improves the efficiency of project safety supervision and reduces the work pressure of safety management personnel. The future work will further improve the adaptability of the model to the obscuring situation, optimize the recognition ability of the visionary target, expand the recognition behavior category, and provide all-round technical guarantee for the intelligent management of civil engineering projects.

References

[1] Tiza, M. T. (2022). Sustainability in the civil engineering and construction industry: A review. Journal of Sustainable Construction Materials and Technologies, 7(1), 30-39.

[2] Zavadskas, E. K., Antucheviciene, J., Vilutiene, T., & Adeli, H. (2017). Sustainable decision-making in civil engineering, construction and building technology. Sustainability, 10(1), 14.

[3] Mei, L., & Wang, Q. (2021). Structural optimization in civil engineering: a literature review. Buildings, 11(2), 66.

[4] Hwang, B. G., Zhu, L., & Ming, J. T. T. (2017). Factors affecting productivity in green building construction projects: The case of Singapore. Journal of Management in Engineering, 33(3), 04016052.

[5] Lee, J., Park, M., Lee, H. S., Kim, T., Kim, S., & Hyun, H. (2017). Workflow dependency approach for modular building construction manufacturing process using Dependency Structure Matrix (DSM). KSCE journal of civil engineering, 21, 1525-1535.

[6] Radujković, M., & Sjekavica, M. (2017). Project management success factors. Procedia engineering, 196, 607-615.

[7] Prascevic, N., & Prascevic, Z. (2017). Application of fuzzy AHP for ranking and selection of alternatives in construction project management. Journal of civil engineering and management, 23(8), 1123-1135.

[8] Mavi, R. K., & Standing, C. (2018). Critical success factors of sustainable project management in construction: A fuzzy DEMATEL-ANP approach. Journal of cleaner production, 194, 751-765.

[9] Fobiri, G., Musonda, I., & Muleya, F. (2022). Reality capture in construction project management: A review of opportunities and challenges. Buildings, 12(9), 1381.

[10] Demirkesen, S., & Ozorhon, B. (2017). Impact of integration management on construction project management performance. International journal of project management, 35(8), 1639-1654.

[11] Pan, Y., & Zhang, L. (2021). Roles of artificial intelligence in construction engineering and management: A critical review and future trends. Automation in Construction, 122, 103517.

[12] Huang, Y., Shi, Q., Zuo, J., Pena-Mora, F., & Chen, J. (2021). Research Status and Challenges of Data-Driven Construction Project Management in the Big Data Context. Advances in Civil Engineering, 2021(1), 6674980.
[13] Al-Zwainy, F. M., Mohammed, I. A., & Al-Shaikhli, K. A. (2017). Diagnostic and assessment benefits and barriers of BIM in construction project management. Civil Engineering Journal, 3(1), 63-77.