scientific reports

OPEN

A multi-object detection method for building fire warnings through artificial intelligence generated content

Jingwei Fu[1], Zhen Xu[1✉], Qingrui Yue[1], Jiarui Lin[2], Ning Zhang[1], Yujie Zhao[1] & Donglian Gu[1]

Timely fire warnings are crucial for minimizing casualties during building fires. In this paper, a multi-object detection method through artificial intelligence generated content (AIGC) is proposed to improve building fire warning capability. First, an AIGC workflow of dataset construction on building fire images is designed, to overcome the limitation due to a serious lack of real building fire images. Validation experiments demonstrate that the detection accuracy of the model trained on the AIGC dataset is only 1.6% lower than that of the model trained on the real image dataset. Subsequently, a multi-object detection model is developed to enhance its feature capture capability, by incorporating the MLCA mechanism into its backbone and replacing the feature fusion layer in its neck. The developed model can detect the flame and smoke of building fires with an accuracy of 95.7%. Finally, the case study involving three real fire incidents demonstrates that the proposed method can detect fires within 2s since the fire starting, which achieves an improvement of at least 6.5 times in the fire warning efficiency compared to the traditional fire alarms. Therefore, the proposed method can deliver timely fire warnings for the evacuation and rescue efforts during building fires.

Building fires may cause casualties and extensive damage due to the rapid spread and difficulty in extinguishment[1,2]. A statistical report from the United States shows that 74% of all building fire-related deaths are attributable to the lack of timely fire warnings[3]. To prevent the consequences caused by building fires, timely fire warning enables the rapid activation of fire suppression systems and ensures sufficient time for personnel evacuation. Therefore, timely fire warnings are critical for minimizing casualties in building fires.

Fire detection methods are employed to achieve timely fire warnings and minimize fire-related damage. Traditional fire detection methods predominantly rely on flame, gas, and temperature sensors, which are typically mounted on the ceiling to maximize their coverage area and detection capability[4,5]. These methods result in detection delays due to the time required for temperature or smoke to reach the sensors[6,7], which increases the response time of the fire warning. Such delays can affect evacuation and the timely activation of fire suppression systems, reducing the effectiveness of fire warnings[8]. Therefore, more efficient detection methods are needed for timely fire warnings.

To overcome the limitations of traditional fire detection methods, scholars have increasingly focused on developing deep learning-based fire detection techniques[9–11]. This approach leverages indoor security cameras for fire detection, enabling more timely and direct fire warnings by utilizing surveillance video[12]. The proposed method is closely integrated with these indoor security cameras. Indoor security cameras installed in the building capture real-time video to obtain information about the current scene. This information is essential for the deep learning model to accurately extract fire-related features. The proposed method can perform real-time analysis based on the latest captured features and quickly issue alarms when fire features are detected.

Deep learning-based fire detection technology relies on high-quality datasets[2] for effective model training[13], as both the quality and quantity of the dataset are crucial to ensuring the accuracy of the model's detection[14]. However, due to the dangerous nature of fire incidents, it is extremely difficult to obtain real fire images, resulting in a serious lack of real image data in existing model training. Given these issues, artificial intelligence-generated content (AIGC) technologies provide a viable solution by generating fire images for model training. AIGC technologies have been widely applied in fields such as text, image, audio, and code generation[15,16]. For instance, DALL-E and Midjourney, as advanced text-to-image technologies[17], offer methods for generating synthetic

[1]Research Institute of Urbanization and Urban Safety, School of Future Cities, University of Science and Technology Beijing, Beijing 100083, China. [2]Department of Civil Engineering, Tsinghua University, Beijing 100084, China. ✉email: xuzhen@ustb.edu.cn

1

images[18,19]. In addition, Generative Adversarial Networks (GANs), such as StackGAN and StyleGAN, have shown the capability to produce realistic and diverse synthetic images[20,21]. Therefore, generative artificial intelligence technology offers a method to expand the dataset with real images, thereby improving the performance of the detection model.

To leverage indoor security cameras for the detection of building fires, the following two challenges need to be addressed:

(1) The scarcity of real fire image samples. After a fire starts, the environment at the scene is complex and it is difficult to obtain a large number of real images of the fire scene. This limitation hinders the development of robust detection models. Employing generative AI methods to construct effective image datasets can augment dataset diversity, thereby facilitating the training of more accurate and reliable detection models.

(2) The efficiency of detection models based on deep learning is insufficient to meet the demands of timely building fire warnings. Building fire detection requires accurate detection of flame or smoke in surveillance images while ensuring detection efficiency to provide accurate fire warnings in various scenarios. Therefore, a detection method for building fires must be developed based on the existing deep-learning models to enable real-time and timely fire warnings in building fires.

For challenge (1), the current method for obtaining training datasets involves using screenshots from indoor security cameras and downloading images online as primary data sources[22]. This approach is limited by insufficient data diversity and inconsistent data quality. Chen collected a multi-mode video dataset using drones[23]. While this method can capture real flame images, it faces challenges in ensuring image quality and clarity due to environmental interference and camera limitations. Furthermore, attempts to augment the dataset sample size through horizontal, vertical, and random flips often lead to excessive similarity among samples, which directly impacts model training accuracy[24]. In addition, researchers often collect fire images from search engines such as Google[25–27]. However, high similarity in image backgrounds reduces sample diversity[28]. Therefore, a dataset construction method is needed that can generate large images to provide training samples for the fire detection model, thereby improving the detection performance of detection models.

For challenge (2), deep learning-based fire detection has achieved a direct mapping from data input to fire detection results, ensuring the reliability and accuracy of video-based detection methods in practical applications[29]. The YOLO (You Only Look Once) network model is widely utilized in numerous real-time object detection applications due to its simplicity, efficiency, and adaptability[30]. The YOLO models have evolved through various versions up to the YOLOv10 model[31]. However, these higher versions, specifically the YOLOv9 and YOLOv10 models, have high computational complexity and resource demands, making them unsuitable for large-scale video detection tasks[32–34]. Numerous studies indicate that the YOLOv8 model is highly adaptable in large-scale fire detection tasks[35]. The object detection speed and efficiency of the YOLOv8 model are better aligned with the requirements of building fire detection compared to the YOLOv9 and YOLOv10 models. Despite these improvements, existing fire detection models still require further enhancement in terms of accuracy and efficiency.

To address the above issues, a multi-object detection method through AIGC is proposed to improve building fire warning capability. First, a fire image generation workflow is designed using Midjourney software, where fire-related keywords are extracted to generate diverse fire images. The detection accuracy of the model trained on the AIGC dataset is compared against that of the model trained on the real image dataset to evaluate whether the AIGC dataset is effective. Next, the MLCA mechanism is introduced to enhance feature detection, and the feature fusion layer is replaced to improve the model's detection efficiency and accuracy. The multi-object detection model is evaluated through performance comparison and ablation experiments. Finally, three cases are detected to demonstrate the method's efficiency for timely fire warnings. The outcomes of this study can be used to offer timely fire warnings, thereby enhancing personnel evacuation and rescue efforts in building fires.
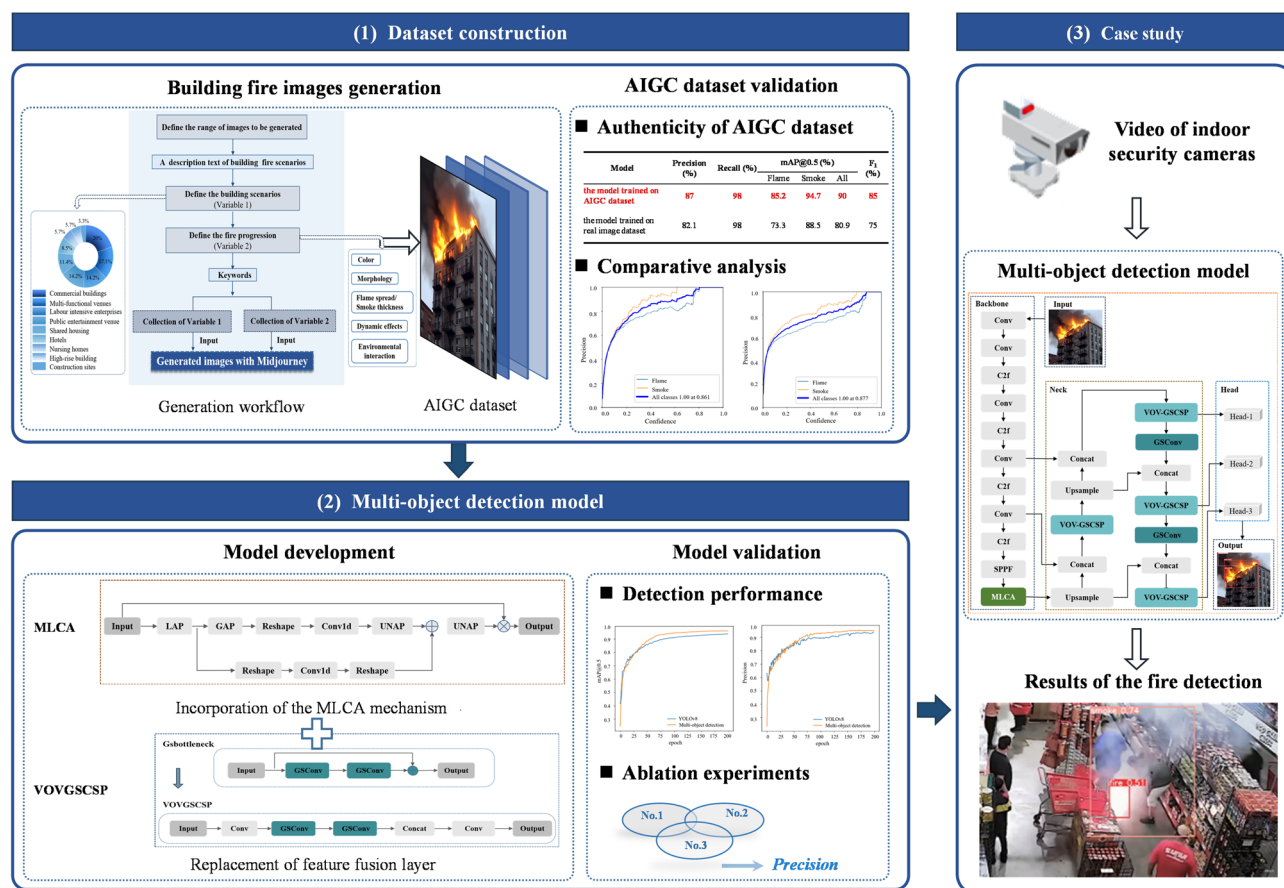
Framework
The research framework of this study is shown in Fig. 1, which consists of three components as follows:

(1) Dataset construction. First, a fire image generation workflow is designed using Midjourney software, where fire-related keywords are extracted and used to generate diverse fire images. Secondly, the validation of the dataset constructed using AIGC demonstrates that the AIGC-based method can expand the number of samples in different fire scenarios.

(2) Multi-object detection model. The MLCA mechanism is introduced, and the feature fusion layers are replaced to enhance the feature fusion capability. Subsequently, detection performance comparison and ablation experiments are performed to demonstrate the effectiveness of the multi-object detection model.

(3) Case study. Three cases are analyzed to verify the effectiveness of the proposed method. The study also demonstrates that the model can effectively detect fires using video captured by indoor security cameras, highlighting its practical applicability in real scenarios.

Dataset construction through AIGC
Construction workflow of the fire image dataset
In the process of building fire evolution, fire causes dynamic damage to the building structure, and this process not only affects the spread of fire, but also significantly changes the main features in the image, so a detailed description of the dynamic damage process can be used in the construction of fire image datasets is of great significance[36,37]. Building fire accidents are characterized by a sudden and dangerous nature; thus, acquiring images of real fire scenes poses a significant safety risk. When a fire incident occurs, factors such as the type and

Fig. 1. The framework of this study.

size of the building, the location of the fire source, and the building materials affect the quality of the image. Therefore, AIGC effectively expands the dataset by synthesizing fire images, thus providing enough training data to help optimize the detection model performance, which solves the problem of scarce fire image data.

In this paper, a construction workflow for the building fire image dataset based on Midjourney V5.2[39] is designed to provide sufficient sample data for the training of fire detection models. Figure 2 shows the workflow for constructing the AIGC-based fire image dataset. First, the range of the generated images must be clarified by specifying fire scenarios and detailed requirements. Then, a descriptive text is created to comprehensively describe different building fire scenarios based on these requirements. The descriptive text consists of two key variables: Variable 1, representing the different fire scenarios, and Variable 2, representing dynamic fire features (e.g., changes in the morphology of the flames and smoke). The combination of variable 1 and variable 2 creates diverse inputs, which are processed by Midjourney to produce a series of building fire scene images. In selecting Variable 1, nine high-risk fire scenarios were chosen to define the environment in which fires occur, taking into account the frequency and impact of building fires. These high-risk scenarios were identified based on factors such as fire incident frequency, building type, and fire impact. For Variable 2, the evolution and morphological characteristics of flames and smoke were refined into 18 categories across five dimensions to capture their dynamics, as summarized in Table 1. In Midjourney, keyword groups are formed by combining the above variables and nesting them within {}. These keyword groups are arranged and combined according to predefined rules, enabling the batch generation of building fire images that meet the target requirements. This workflow not only ensures high realism and diversity in the generated fire images but also provides more accurate training data for the fire detection model.

Through the workflow designed for this paper, 2000 building fire images were generated within 8 h, with the images being clear and accurately reflecting the details of the building fire scene. Figure 3 illustrates the generated fire images. The generated images encompass various fire scenarios, building types, and environmental conditions. Flames are depicted with varying intensities, accompanied by smoke effects that differ in both density and distribution. These details capture the intricate characteristics of fire incidents, reflecting the diversity inherent in real fire scenarios. Such variations are crucial for training detection models. This dataset serves as an invaluable resource for training fire detection models, enhancing their ability to detect fire under different lighting and weather conditions.
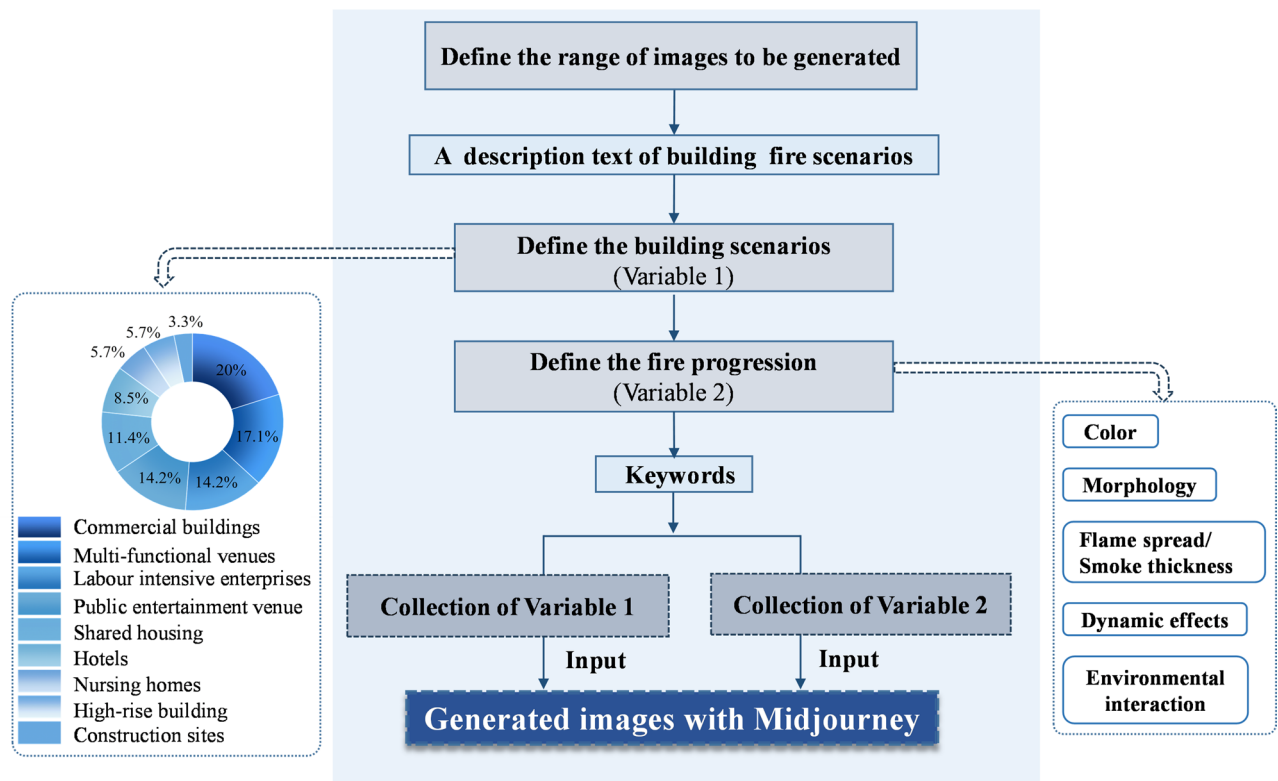
Fig. 2. The workflow for constructing the AIGC-based fire image dataset.

AIGC dataset validation

Validation experiment

The validation experiments aim to determine whether the performance of models trained on the AIGC-generated dataset is consistent with that of models trained on the real fire image dataset. To achieve this, the AIGC-generated dataset and the real image dataset are used to train the models separately. The validation of the AIGC dataset consists of two experiments: evaluating the detection performance of the two models and comparing their performance on the same dataset. For Experiment 1, the 2000 AIGC-generated fire images and 16,800 real fire images were divided into training and validation sets in an 80:20 ratio. Both datasets were used to train the YOLOv5 model, and the models' evaluation metrics were compared. Experiment 2 focused on comparing the detection precision of the two models on a dataset of 1000 real fire images. These two experiments aimed to analyze the impact of different datasets on the performance of fire detection models.

In this study, the experimental platform is built on a Windows 10 operating system with a 12th Gen Intel Core (TM) i7-12700 H CPU and an NVIDIA GeForce RTX 3090 GPU. The platform utilizes Python 3.8[40] as the programming language and PyTorch as the framework. This paper evaluates the model performance in terms of both accuracy and inference speed. Precision measures the proportion of correctly detected objects to all detected objects, recall evaluates the ability of the model to detect actual objects, and F_1 score is used as a combined indicator of precision and recall to assess the overall quality of the model. The inference speed is evaluated by mAP@0.5 (denotes the mAP calculated at an IOU threshold of 0.5) and FPS (frames per second), which is critically used for real-time detection. These metrics provide a comprehensive and scientific basis for different datasets trained model evaluation.

Validation results

The model trained on the AIGC dataset is denoted as the AIGC model, and the model trained on the real dataset is denoted as the real image model.

Experiment 1 compared the evaluation metrics between models trained on the AIGC dataset and those trained on the real image dataset. The results are detailed in Table 2, which demonstrates the superior performance of the AIGC model in comparison to the real image model. Specifically, both models achieve a recall of 98.0%, indicating that the two models are equally effective in detecting fire characteristics. However, the AIGC model achieves 4.9% higher precision than the real image model. This higher precision suggests that the AIGC model generates fewer false positives, thereby improving the reliability of fire warnings. In addition, a comparison of the mAP@0.5 reveals the advantage of the AIGC model. For mAP@0.5 (flame), the AIGC model achieves 85.2%, surpassing the real image model's 73.3% by 11.9%. For mAP@0.5 (smoke), the AIGC model achieves 94.7%, outperforming the real image model by 6.2%. Overall, the AIGC model achieves a mAP@0.5 of 90.0%, which is 9.1% higher than the real image model's 80.9%. The F_1 score of the models further highlights the advantages of the AIGC model. It achieves 85%, exceeding the 75.0% of the image dataset model. Despite being trained on

| Fire progression	Flame		Smoke	
	Keyword	Explanation	Keyword	Explanation
Color	Bright orange-red flame	Typical flame color	Thick black smoke	Indicates burning oil, rubber, or plastic with a violent fire
	Vivid crimson flame	Indicates intense and hazardous flame, showing aggressive burning	Pale grayish-white smoke	Indicates light burning
	Luminous blue flame	Indicates high-temperature combustion	Gray smoke	Indicates burning of common materials with medium intensity
	Orange flame with black smoke	Indicates incomplete combustion	Yellow smoke	Indicates burning chemicals or the presence of toxic gases
Morphology	Moving flame	Flame sways and moves upwards, often indicating a weak fire	Smoke rising upward	Smoke rises vigorously, often indicating intense combustion
	Rising flame	Displays the presence of smoke, indicating active combustion	Smoke spreading in the air	Smoke spreads and disperses in the air, covering a large area
	Pointed flame	Flame takes on a sharp shape	Smoke forming vortex shapes	Smoke forms vortex or spiral shapes, possibly due to wind or airflow effects
	Wave-like flame	Flame exhibits a wave-like pattern, influenced by airflow	Floating smoke belt	Smoke forms a floating belt, moving with the airflow, indicating a specific flow pattern
Flame spread/ Smoke thickness	Intense flame	Indicates strong fire intensity	Dense smoke	Often seen in severe fire situations, with low visibility
	Broad, expansive flame	Flame spreads in multiple directions	Thin smoke	Visible and relatively light, common in the early or late stages of a fire
	Weak flame	Indicates an initial or fading stage of combustion	Thick smoke layer	Often found indoors or in windless environments
Dynamic effects	Rapidly flickering flame	Indicates characteristics under wind or turbulent conditions	Slowly rising smoke	Smoke rises slowly and steadily
	Slowly burning flame	Flame burns steadily	Rapidly expanding smoke	Indicates fire intensification and an urgent, hazardous scene
	Ejecting flame	Resembles explosive or jet-like combustion	Swirling rising smoke	Typically influenced by airflow or flame movement
Environmental interaction	Flame engulfing building	Demonstrates the intensity and destructive force of the fire	Smoke surrounds the building	Smoke spreads around the building
	Flame hidden in smoke	Flame partially visible within dense smoke	Smoke conceals the sky	Smoke covers the sky
	Flame illuminating night	In a night environment, due to the presence of flame that illuminate the surroundings	Smoke interwoven with flame	Smoke and flame intertwine
	Flame interacting with wind	Displays a dynamic interaction between flame and surrounding airflow	Smoke dispersed by wind	Displays the impact of environmental factors on the fire scene

Table 1. Keywords and explanations for Variable 2.

Fig. 3. The generated images of building fires based on the construction workflow in this study.

Model	Precision (%)	Recall (%)	mAP@0.5 (%)			F$_1$ (%)
			Flame	Smoke	All	
AIGC model	87.0	98.0	85.2	94.7	90.0	85.0
Real image model	82.1	98.0	73.3	88.5	80.9	75.0

Table 2. Evaluation metrics between the models trained on the AIGC and the real image datasets.

(a) Precision of the AIGC model.   (b) Precision of the real image model.
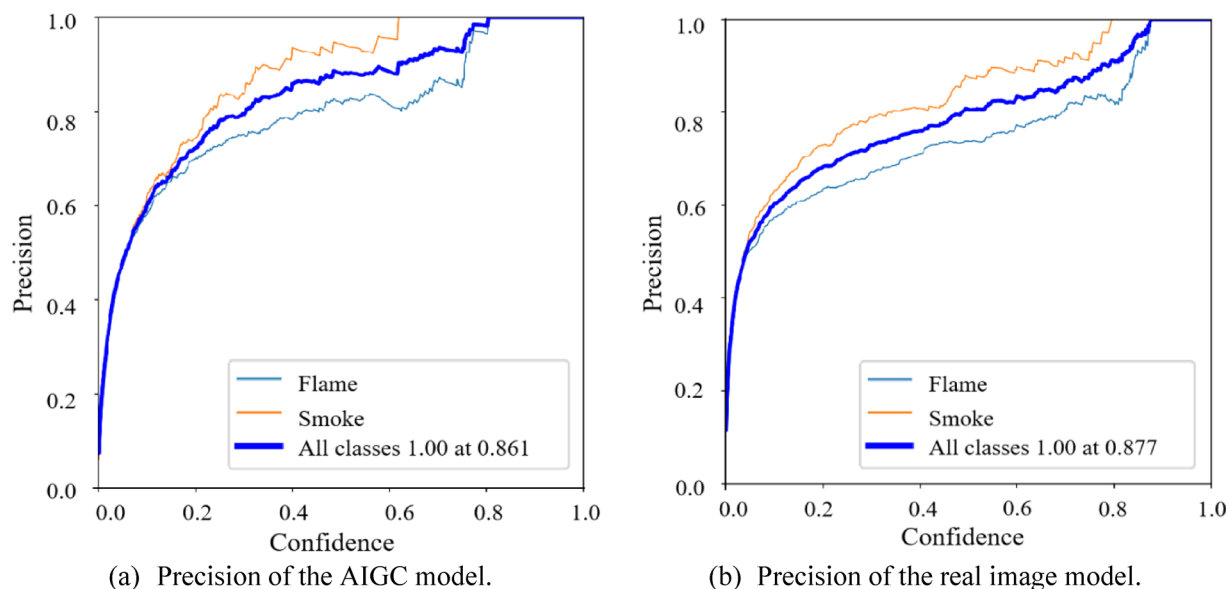
Fig. 4. Precision comparison of the AIGC and the real image models on the same image dataset.

a smaller dataset, the AIGC model consistently outperforms the real image model across various evaluation metrics, indicating that AIGC-generated data can effectively complement real image datasets in training fire detection models.

Experiment 2 compares the precision of both models on the same image dataset. Figure 4 shows precision of 86.1% and 87.7%, respectively, on the same image dataset. Theexperiments show that the AIGC model has a precision of only 1.6% lower than the realimage model, demonstrating the effectiveness of the AIGC dataset in detecting real fire. Italso demonstrates the reliability of the AIGC-generated dataset. The performance gap between the AIGC model and the real image model in Experiment 2 is minimal, furtherdemonstrating the effectiveness and reliability of the AIGC model for real fire detection tasks.

In summary, the results of these validation experiments show that the AIGC model outperforms the real image model in key evaluation metrics. Although the AIGC model was trained on a smaller dataset, it demonstrated greater precision and reliability in detecting real fire images. The results from both experiments highlight the significant potential of AIGC-generated datasets to improve the performance of fire detection models, demonstrating their ability to complement real image datasets.

Multi-object detection model
Model development
Applying the YOLOv8 model in building fire scenarios still faces several challenges despite it demonstrating high efficiency and accuracy in detecting flames and smoke. First, complex backgrounds and varying lighting conditions in building fire scenes pose challenges to feature extraction and object detection. Second, the performance of the YOLOv8 model in detecting objects of varying scales, particularly smaller flames and fine smoke, requires further enhancement. Moreover, the high computational complexity of the YOLOv8 model may hinder its application in real-time surveillance systems.

To address these challenges and improve the detection performance of the detection model in building fire scenarios, this paper develops a multi-object detection model based on the YOLOv8 model. The approach aims to improve the model's ability to detect flames and smoke through structural modifications and algorithm optimization. The overall architecture of the multi-object detection model is presented in Fig. 5, and the architectures of the sub-modules are shown in Fig. 6.

In the backbone sub-module, the fire image is performed with a series of convolutions to output fire feature information. Simultaneously, the attention mechanism is introduced at the 10th layer of the model to focus channel and spatial information to enhance the backbone's ability to extract flame and smoke features. The introduction of the attention mechanism can retain channel and spatial information. In recent years, attention
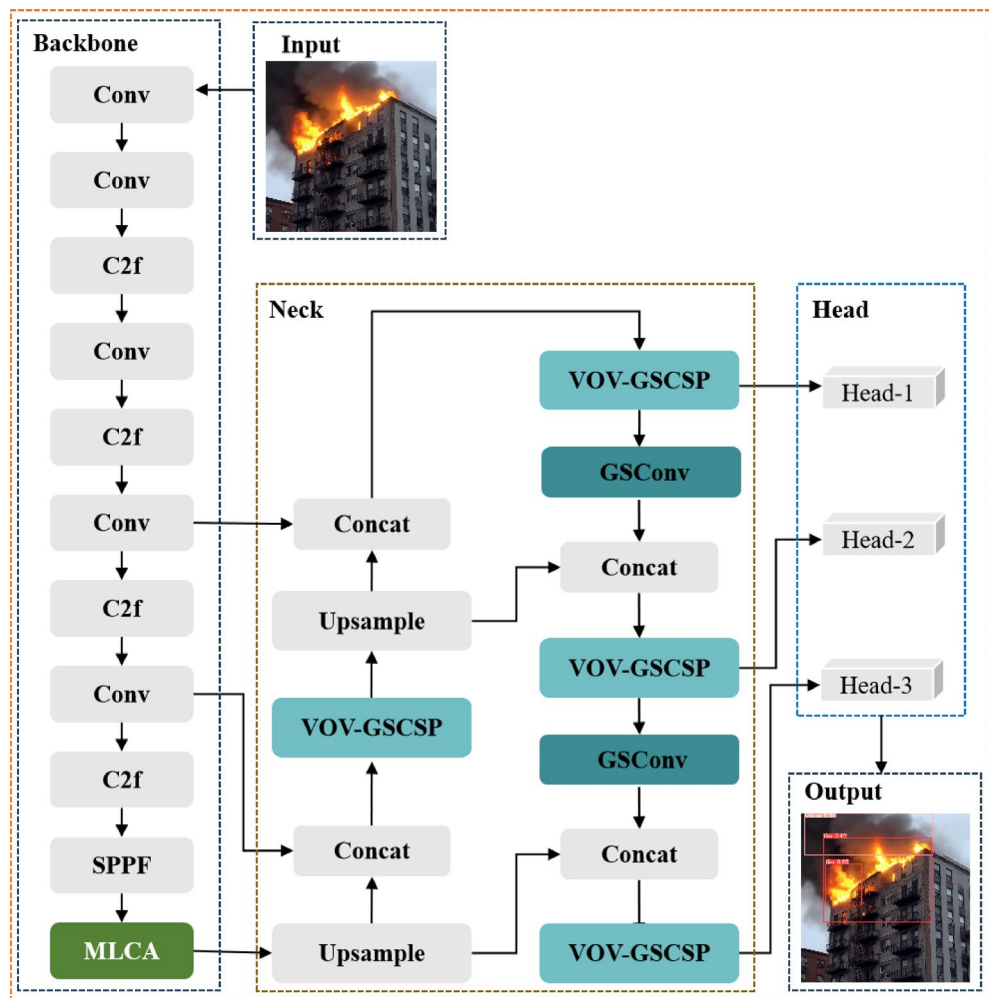
Fig. 5. The overall architecture of the multi-object detection model.
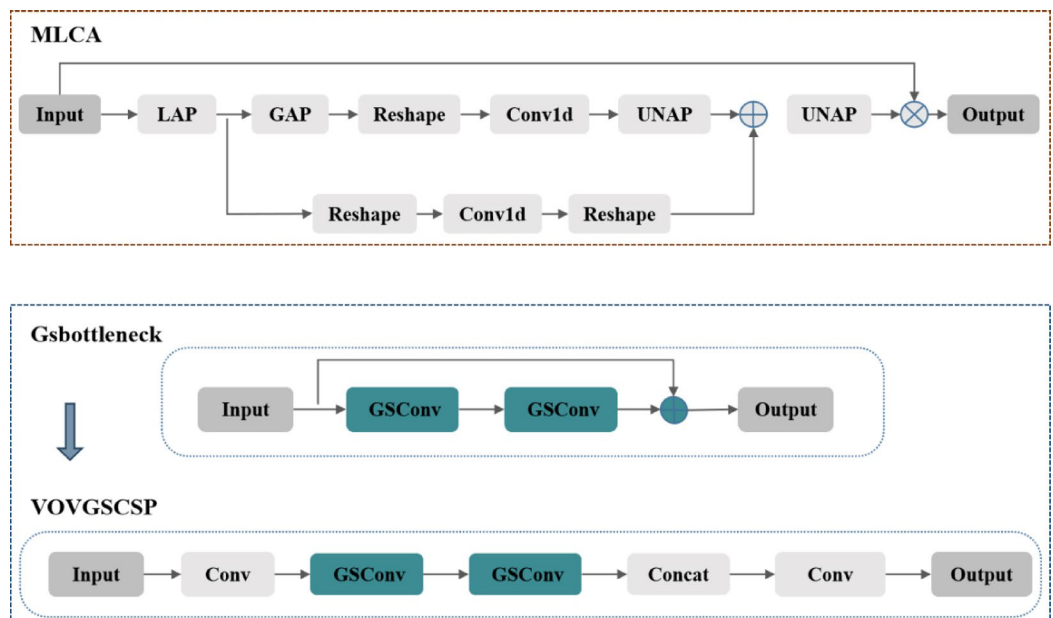

Fig. 6. The sub-module architecture of the multi-object detection model.

mechanisms have been proven effective in many studies for detecting objects[35], such as large separable kernel attention (LSKA)[36], convolutional block attention module (CBAM)[40–44], and coordinate attention (CoordAtt). However, these approaches fall short in terms of feature extraction, computational efficiency, and generalization ability. The Mixed Local Channel Attention (MLCA) mechanism boosts performance by reducing information reduction and magnifying global interactive representations, especially in complex environments. Therefore, the MLCA mechanism is introduced into the backbone of the model in this paper to capture important information in images and enhance the model's feature representation capabilities. By incorporating the MLCA mechanism into its backbone, the precision increases by 2.2%, and mAP@0.5 improves by 1.7% compared to the YOLOv8 model, demonstrating the significant impact of the MLCA mechanism on feature extraction. Additionally, the C2f and Conv convolutions in the neck of the YOLOv8 model are replaced with VOV-GSCSP and GSConv, thereby reducing FLOPs. The neck layer of the multi-object model balances complexity and accuracy, achieving higher computational efficiency while effectively improving detection accuracy for flame and smoke. The neck module contains mainly sub-modules for concat, upsampling, VOV-GSCSP, and GSConv. The VOV-GSCSP allows parameters to be fused from different backbone layers to different detection layers, greatly improving the feature fusion capability of the network. By incorporating the MLCA mechanism into its backbone and replacing the feature fusion layer in its neck, the precision reaches 95.7%, and the mAP@0.5 reaches 96.4% compared to the YOLOv8 model.

Model validation

Two experiments were used to evaluate the performance of the developed multi-object detection model: the detection performance of the multi-object detection model and an ablation experiment to assess the impact of the sub-module architecture.

Detection performance

Figure 7 presents the change curves of evaluation metrics mAP@0.5 and precision for the original YOLOv8 model and the multi-object detection model. In Fig. 7a, the mAP@0.5 curve shows the overall detection accuracy of both models. The multi-object detection model outperforms YOLOv8 during the early training phase. It maintains high mAP@0.5 values across all iterations and converges faster initially, indicating that the developed model learns flame and smoke features more quickly and comprehensively. In Fig. 7b, compared to the YOLOv8 model, the multi-object detection model exhibits higher accuracy throughout the training process, indicating a reduction in the false alarm rate. This highlights the greater robustness and accuracy of the developed model when dealing with complex fire detection scenarios.

Evaluation metrics between the YOLOv8 model and the multi-object detection model are presented in Table 3. Although the FPS of the multi-object detection model is lower than that of the YOLOv8 model, it still surpasses the frame processing capability of indoor security cameras. This indicates that the model is suitable for timely fire warnings in building fires. In addition, the multi-object detection model shows significant improvement in detection accuracy on all metrics. The improvement in the F_1 score further highlights the balance between precision and recall, ensuring reliable detection performance. These results indicate that the multi-object detection model successfully balances detection accuracy and computational efficiency. This balance of accuracy and efficiency ensures the applicability of the model in building fire detection.

(a) Comparison of mAP@0.5

(b) Comparison of precision
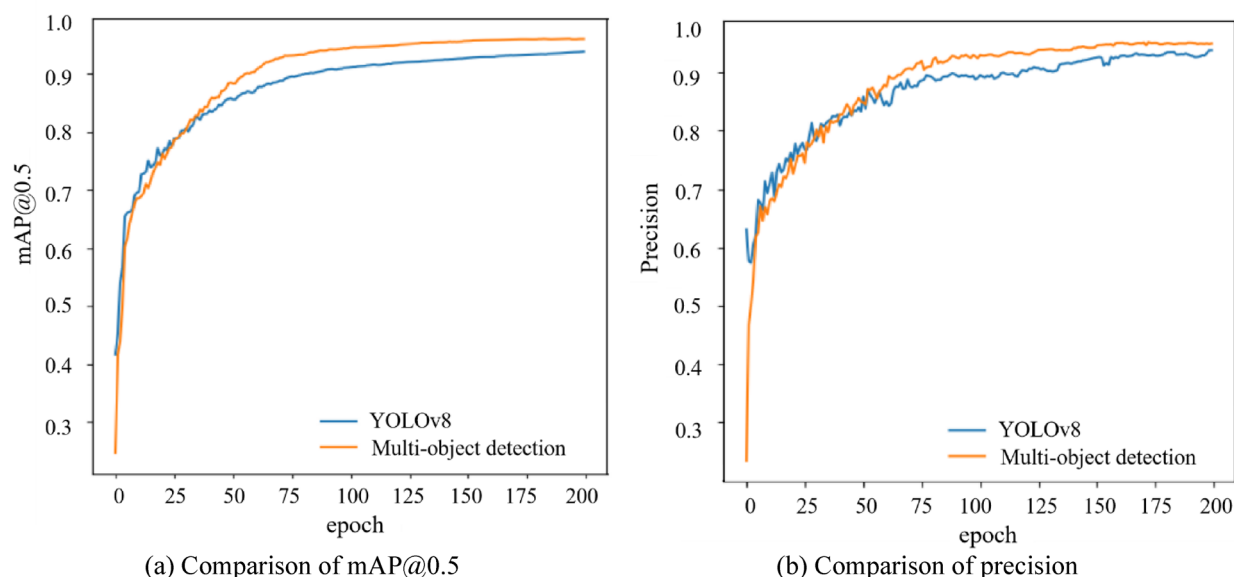
Fig. 7. The comparison of mAP@0.5 and precision between the YOLOv8 and multi-object detection model.

| Model | Precision (%) | Recall (%) | mAP@0.5 (%) | | | F_1 (%) | FPS (f/s) |
			Flame	Smoke	All		
YOLOv8	92.6	97.0	90.6	97.1	93.8	89.0	228
Multi-object detection	95.7	98.0	95.1	97.7	96.4	93.0	178

Table 3. Evaluation metrics comparison between the YOLOv8 and the multi-object detection model.

| Experiment Number | Improvement | | Performance metrics for fire detection | | | |
| | MLCA | Replacement of feature fusion layer | | mAP@0.5 (%) | | |
			Precision (%)	Flame	Smoke	All
I	×	×	92.6	90.6	97.1	93.8
II	√	×	94.8	93.1	98.0	95.5
III	√	√	95.7	95.1	97.7	96.4

Table 4. Ablation experiments of the multi-object detection model.

Case	Detection challenges	Case details	Video screenshot
1	Complex environment +Transient	"9-29" Fire incident in Ningbo, China	
2	Complex environment +Ignition of small target	Fireworks ignited at a Supermarket, in the United States	
3	Low clarity +Ignition of small target +Transient	"12-15" Fire incident in Henan Province, China	

Table 5. Introduction of three actual fire cases.

Ablation experiments

The ablation experiments were conducted to validate the effect of the improved part of the model on the fire recognition effect. The ablation experiment consisted of 3 experiments: Experiment I used the original YOLOv8 model, Experiment II introduced the MLCA mechanism into its backbone, and Experiment III introduced the MLCA mechanism into its backbone and replaced the feature fusion layer in its neck. By comparing the results of these experiments, the specific contribution of each improvement to the model performance can be explicitly assessed. The ablation experimental results of the multi-object detection model are shown in Table 4. Comparing the results of Experiment I and Experiment II, the precision increases by 2.2%, and mAP@0.5 improves by 1.7%. Comparing the results of Experiment I and Experiment III, the precision reaches 95.7%, and the mAP@0.5 (All) achieves the highest value of 96.4%. Experiment III confirms that the model developed in this study is more capable of acquiring fire characteristic information. The results of the ablation experiments confirm that the multi-object detection model has superior detection performance.

Case study
Case introduction

In this study, three cases, consisting of news reports and surveillance videos, were selected to detect fire incidents to evaluate the performance and robustness of the proposed method, the case detection challenges, and the details listed in Table 5.

The surveillance video for Case 1 was taken from post-disaster news reports, and the surveillance video recorded the course of a major fire incident in Ningbo, China, in 2019, where a fire broke out in a warehouse for daily necessities. The surveillance video from news reports reveals the complexity of the environment in which

the fire started and the transient nature of the fire. However, due to the smoke concentration in the early stages of the fire being below the activation threshold of the smoke alarm, the fire alarm was delayed for 20 s.

The surveillance video for Case 2 captured a fireworks ignition incident at a supermarket in the United States. At the time of the fire's outbreak, customers were selecting items near a shelf. However, due to the unique chemical properties of the fireworks, the fire spread rapidly. Furthermore, the smoke alarm was not triggered until 13 s after the fire ignited.

The surveillance video for Case 3 captured the fire incident in Henan Province, China, in 2014, characterized by an unclear visual representation and transient. However, the smoke concentration at the start of the fire had not yet reached the activation threshold of the smoke alarm, and the fire alarm was only issued 7 s after the fire broke out. Both Case 2 and Case 3 were derived from surveillance video captured by indoor security cameras, and the blurred images present challenges for detection.

Results of the multi-object detection

To further analyze the process of fire detection, this section presented an analysis of different stages of fire images. The increase in video complexity reduced the FPS to 52–71, which was still suitable for video fire detection applications. As seen from Table 6, the model achieved optimal performance in Case 1 due to the clarity of the video. In Case 2 and Case 3, the relatively blurry case video resulted in a slight increase in detection latency.

In terms of smoke detection, Table 6 showed the results of early fire detection, indicating the effectiveness of detecting smoke at an early stage. This early detection was crucial for preventive action and emphasized the sensitivity of the multi-object detection model to the smoke characteristic. In terms of flame detection, the multi-object model successfully detected flames in both localized and intense stages. The model captured the distinctive characteristics of flames at the initial stage and the stage of intense combustion. This suggested that the model could detect flames at all stages of fire development, providing timely warnings in the early stages of a fire.

In conclusion, the accuracy of the detection and the high FPS proved the validity of the multi-object detection model and the reliability of fire monitoring, ensuring that it met the requirements for effective video fire detection.

The fire detection time results are shown in Table 7. In three cases, the detection time was maintained under 2 s, ensuring that the model provided timely warnings during the early stages of a fire. In Case 1, the fire scene included multiple obstacles, but the multi-object detection model could detect the fire within 2 s of flame. In Case 2, despite the low clarity smoke characteristics, smoke was detected 2 s after the fire outbreak. In Case 3, despite the low clarity of the video, the model detected smoke immediately after the fire started, demonstrating the robustness of the detection model.

The warning time for the traditional fire alarms (e.g. smoke sensors) and the developed multi-object detection model were compared shown in Fig. 8. In the above three cases, the time required by the traditional fire alarms was 20, 13, and 7 s, respectively, whereas the warning time by the developed multi-object detection model was significantly shorter, taking only 2, 2, and 1 s, respectively. These results demonstrated the substantial performance advantages of the multi-object detection model over traditional fire alarms, particularly in terms of detection efficiency. In Case 1, the model detected fires in 2 s, compared to the 20 s required by the traditional method, achieving 10 times in efficiency. This highlighted the model's ability to efficiently identify early fire characteristics, even in complex environments. In Case 2, the developed model made an improvement of 6.5 times in the efficiency of fire warnings. Similarly, in Case 3, the proposed method also improved 7 times in efficiency. This demonstrated that the proposed method was at least 6.5 times more efficient on fire warnings compared to traditional fire alarms in the above cases. In summary, the developed multi-object detection model enabled the timely detection of fire characteristics in building fires.
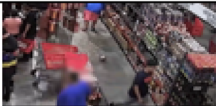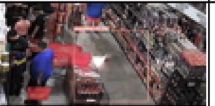
Case	FPS (f/s)	Smoke detection	Flame detection	
			Localized flame	Intense flame
1	71			
2	52			
3	66			

Table 6. Detection results for flame and smoke in building fires in three actual fire cases.

Case	Fire time (s)		Detection time (s)		Detection result	
	Video screenshot	Start time	Video screenshot	Detection time	Time	Object
1		12:50:04		12:50:06	2s	Flame
2		0th second		2nd second	2s	Smoke
3		00:25:19		00:25:20	1s	Smoke

Table 7. Fire detection time by the proposed method in three actual fire cases.
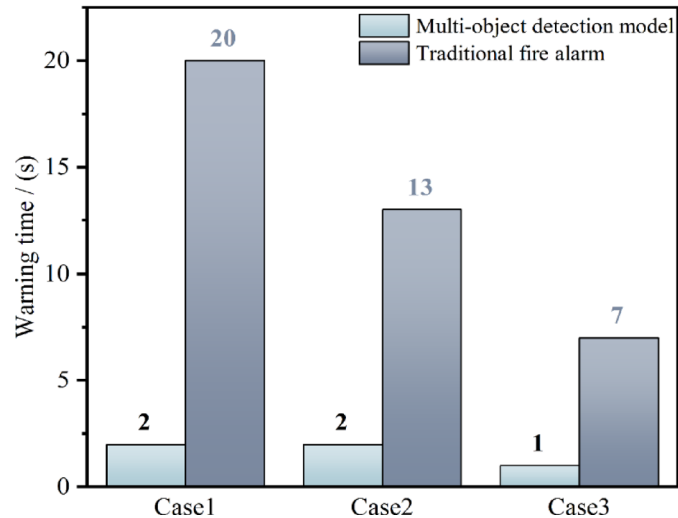
Fig. 8. Warning time comparison between the multi-object detection model and the traditional fire alarms in three actual fire cases.

Conclusion

In this paper, a multi-object detection method through AIGC is proposed. The effectiveness of the proposed method is demonstrated through its application to three historical fire videos. Some conclusions are summarized as follows:

(1) For the AIGC-dataset, the AIGC-generated fire images can be used to expand the dataset of building fire images and address the limitation caused by a serious shortage of real building fire images. The validation experiments indicate that the detection accuracy of the model trained on AIGC-generated fire images is only 1.6% lower than that of the model trained on the real image dataset.

(2) For the detection model, the developed model enhances the detection of flames and smoke in complex environments, enabling more timely and accurate fire warnings. The developed multi-object detection model achieves a mAP@0.5 of 96.4%, representing a 2.6% increase compared to the YOLOv8 model. Additionally, its precision reaches 95.7%, which is 3.1% higher than that of the YOLOv8 model.

(3) For the method, the three case studies demonstrate that the proposed detection method can accurately detect building fires within 2 s. The proposed method shows a significant advantage over traditional fire warning methods, being 10 times, 6.5 times, and 7 times more efficient in the three cases, respectively.

(4) For application, the method proposed in this paper leverages indoor security cameras for fire detection, enabling more timely and direct fire warnings. It can provide building occupants with more time to evacuate, significantly reducing the risk of fire-related casualties.

The proposed method is limited to data captured by indoor security cameras, which restricts its fire detection scope primarily to indoor spaces. If outdoor cameras were added, it would be possible to train the model using the AIGC-generated dataset of outdoor images to detect outdoor fires. However, this aspect has not been explored in the present study. In the future, traditional fire detection methods can also be combined with the proposed method for fire detection to improve the efficiency of fire detection.

Data availability
The data are available from the corresponding author on reasonable request.

References
1. Yar, H., Khan, Z. A., Ullah, F. U. M., Ullah, W. & Baik, S. W. A modified YOLOv5 architecture for efficient fire detection in smart cities. *Expert Syst. Appl.* **231**, 120465 (2023).
2. Avazov, K., Mukhiddinov, M., Makhmudov, F. & Cho, Y. I. Fire detection method in smart City environments using a Deep-Learning-Based approach. *Electronics* **11**, 73 (2021).
3. Ahn, Y., Choi, H. & Kim, B. S. Development of early fire detection model for buildings using computer vision-based CCTV. *J. Building Eng.* **65**, 105647 (2023).
4. Wen, Z., Xie, L., Feng, H. & Tan, Y. Robust fusion algorithm based on RBF neural network with TS fuzzy model and its application to infrared flame detection problem. *Appl. Soft Comput.* **76**, 251–264 (2019).
5. Jadon, A. et al. A specialized lightweight fire & smoke detection model for real-time IoT applications. (2019).
6. Talaat, F. M. & ZainEldin, H. An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Comput. Appl.* **35**, 20939–20954 (2023).
7. Tian, Y., Ren, J., Xu, Z. & Qi, M. A Cost–Benefit analysis framework for City-Scale seismic retrofitting scheme of buildings. *Buildings.* **13**, 477 (2023).
8. Zhao, H., Jin, J., Liu, Y., Guo, Y. & Shen, Y. FSDF: A high-performance fire detection framework. *Expert Syst. Appl.* **238**, 121665 (2024).
9. Gao, S. et al. Two-stage deep learning-based video image recognition of early fires in heritage buildings. *Eng. Appl. Artif. Intell.* **129**, 107598 (2024).
10. Bu, F. & Gharajeh, M. S. Intelligent and vision-based fire detection systems: A survey. *Image Vis. Comput.* **91**, 103803 (2019).
11. Wang, Z., Zhang, T., Wu, X. & Huang, X. Predicting transient Building fire based on external smoke images and deep learning. *J. Building Eng.* **47**, 103823 (2022).
12. Reddy, P. D. K., Margala, M., Shankar, S. S. & Chakrabarti, P. Early fire danger monitoring system in smart cities using optimization-based deep learning techniques with artificial intelligence. *J. Reliable Intell. Environ.* **10**, 197–210 (2024).
13. Atitallah, S. B., Driss, M., Boulila, W. & Ghézala, H. B. Leveraging deep learning and IoT big data analytics to support the smart cities development: review and future directions. *Comput. Sci. Rev.* **38**, 100303 (2020).
14. Wang, Z., Zhang, T. & Huang, X. Predicting real-time fire heat release rate by flame images and deep learning. *Proc. Combust. Inst.* **39**, 4115–4123 (2023).
15. Rapp, A., Di Lodovico, C., Torrielli, F. & Di Caro, L. How do people experience the images created by generative artificial intelligence? An exploration of People's perceptions, appraisals, and emotions related to a Gen-AI text-to-image model and its creations. *Int. J. Hum Comput. Stud..* **193**, 103375 (2025).
16. Lin, H. et al. Comparing AIGC and traditional Idea generation methods: evaluating their impact on creativity in the product design Ideation phase. *Think. Skills Creativity.* **54**, 101649 (2024).
17. Agnese, J., Herrera, J., Tao, H. & Zhu, X. A. Survey and taxonomy of adversarial neural networks for text-to-image synthesis. *WIREs.* (2019).
18. Nichol, A. et al. GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models. http://arxiv.org/abs/2112.10741 (2022).
19. Saharia, C. et al. Photorealistic Text-to-Image diffusion models with deep Language Understanding. *Adv. Neural Inform. Process. Syst.* **35**, (2022).
20. Zhang, H. et al. StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2017).
21. Karras, T. et al. Analyzing and improving the image quality of StyleGAN. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8107–8116 (IEEE, 2020).
22. Hasib, R., Jan, A. & Khan, G. M. Real-time anomaly detection for smart and safe city using spatiotemporal deep learning. In *Proceedings of the 2022 2nd International Conference on Artificial Intelligence (ICAI)*, 79–83 (IEEE, 2022).
23. Chen, X. et al. Wildland fire detection and monitoring using a Drone-Collected RGB/IR image dataset. *IEEE Access.* **10**, 121301–121317 (2022).
24. Wu, Z., Xue, R. & Li, H. Real-time video fire detection via modified YOLOv5 network model. *Fire Technol.* **58**, 2377–2403 (2022).
25. Qin, Y. Y., Cao, J. T. & Ji, X. F. Fire detection method based on depthwise separable Convolution and YOLOv3. *Int. J. Autom. Comput.* **18**, 300–310 (2021).
26. Zhao, L., Zhi, L., Zhao, C. & Zheng, W. Fire-YOLO: A small target object detection method for fire inspection. *Sustainability* **14**, 4930 (2022).
27. Zhao, Y., Zhang, H., Zhang, X. & Chen, X. Fire smoke detection based on target-awareness and depthwise convolutions. *Multimed. Tools Appl.* **80**, 27407–27421 (2021).
28. Li, Y., Wu, A., Dong, N., Han, J. & Lu, Z. Smoke recognition based on deep transfer learning and lightweight network. In *Proceedings of the 2022 2nd International Conference on Artificial Intelligence (ICAI)*, 8617–8621 (IEEE, 2019).
29. Cheng, G. et al. Visual fire detection using deep learning: A survey. *Neurocomputing.* **596**, 127975. https://doi.org/10.1016/j.neucom.2024.127975 (2024).
30. Zou, Z., Chen, K., Shi, Z., Guo, Y. & Ye, J. Object detection in 20 years: a survey. In *Proceedings of the IEEE*, 257–276 (IEEE, 2023).
31. Wang, C. Y. & Liao, H. Y. M. YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems (2024).
32. Qiu, X., Chen, Y., Cai, W., Niu, M. & Li, J. LD-YOLOv10: A lightweight target detection algorithm for drone scenarios based on YOLOv10. *Electronics* **13**, 3269. https://doi.org/10.3390/electronics13163269 (2024).
33. An, R., Zhang, X., Sun, M. & Wang, G. GC-YOLOv9: innovative smart City traffic monitoring solution. *Alex. Eng. J.* **106**, 277–287 (2024).
34. Xu, G., Yue, Q. & Liu, X. Real-time monitoring of concrete crack based on deep learning algorithms and image processing techniques. *Adv. Eng. Inform.* **58**, 102214 (2023).
35. Guo, M. H. et al. Attention mechanisms in computer vision: A survey. *Comp. Visual Media.* **8**, 331–368 (2022).

36. San, B. B., Xu, S. Z., Shan, Z. W., Chen, W. & Looi, D. T.-W. Investigation on probability model of bending moment-rotation relationship for bolt-ball joints. *J. Constr. Steel Res.* **226**, 109193 (2025).
37. Shan, Z., Ma, H., Yu, Z. & Fan, F. Dynamic failure mechanism of single-layer reticulated (SLR) shells with bolt-column (BC) joint. *J. Constr. Steel Res.* **169**, 106042 (2020).
38. Midjourney *Midjourney*. https://www.midjourney.com/website.
39. Welcome (ed) to Python.org. *Python.org*. https://www.python.org/ (2025).
40. Tie, J. et al. LSKA-YOLOv8: A lightweight steel surface defect detection algorithm based on YOLOv8 improvement. *Alex. Eng. J.* **109**, 201–212 (2024).
41. Zamri, N. M. Enhanced small drone detection using optimized YOLOv8 with attention mechanisms. *IEEE Access..* **12**, 90629–90643 (2024).
42. Yang, W. et al. Deformable Convolution and coordinate attention for fast cattle detection. *Comput. Electron. Agric.* **211**, 108006 (2023).
43. Wan, D. et al. Mixed local channel attention for object detection. *Eng. Appl. Artif. Intell.* **123**, 106442 (2023).
44. Zhang, Y. et al. Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. *J. Neurosci. Methods.* **353**, 109098 (2021).

Acknowledgements

Author contributions

J.F., Z.X., and Q.Y. proposed a method; J.F. and Z.X. designed the workflow and developed the detection model; N.Z. and Y.Z. designed the validation experiments; J.F., N.Z., and J.L. assisted with manuscript formatting review; Z.X., Q.Y., J.L., and D.G. guided the experiment and the final revision of the manuscript; Z.X. and D.G. offered financial support. All the authors contributed extensively to the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.