Kalpa Publications in Computing
Volume 22, 2025, Pages 817–828
Proceedings of The Sixth International Conference on Civil and Building Engineering Informatics

Visual-Driven Inspection for Collecting the Status of Fire Safety Equipment Status in Building Spaces

Fangzhou Lin[1], Zhengyi Chen[2], Xiao Zhang[3], Boyu Wang[4], Jack C.P. Cheng[5]

1) Ph.D. Student, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: flinan@connect.ust.hk
2) Ph.D., Postdoctoral Fellow, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: zchenfq@connect.ust.hk
3) Ph.D. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: xiao.zhang@connect.ust.hk
4) Ph.D., Postdoctoral Fellow, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: bwangbb@connect.ust.hk
5) Ph.D., Prof., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: cejcheng@ust.hk

Abstract: Installed fire safety equipment within buildings plays a crucial role in ensuring the safety of personnel and minimizing losses. Nevertheless, if not maintained appropriately, these devices may fail to function optimally in emergency situations. As building sizes continue to grow, traditional manual inspection methods encounter significant challenges, including a heavy workload and complex information recording tasks. To tackle these issues, advanced emergency equipment detection frameworks and improvement plans have been put forward. This framework is specifically designed to overcome the problem of remote inspection being unable to accurately locate objects by establishing spatial relationships among devices, cameras, and trajectories. Firstly, the improved detection algorithm is utilized to detect objects of interest. Subsequently, these objects are located through a tracking algorithm and Visual Simultaneous Localization and Mapping (vSLAM). The on-site experimental results clearly show that the framework can effectively solve various types of equipment detection problems in a wide range of complex scenarios and holds great promise for replacing manual labor.

Keywords: Fire Safety Equipment, Visual Positioning, Object Tracking, Equipment Detection

1. INTRODUCTION

Emergency equipment, designed for crisis response, safeguards individuals during unforeseen events, reducing harm. For instance, fire extinguishers combat fires, exit signs guide evacuations, and alarms prompt action (Xin & Huang, 2013). Diverse buildings require tailored emergency setups, like fire extinguishers, lighting, and communication devices (Ivanov & Chow, 2022). Hospitals, for instance, need medical gear and backup power (Tang, Fitzgerald, Hou, & Wu, 2014). The state of emergency equipment is crucial for overall safety. Inadequate gear or outdated regulations in buildings impair emergency responses, emphasizing the need for well-maintained equipment (Dong, You, & Hu, 2014; Walters & Hastings, 1998). Regular checks are vital to ensure equipment integrity and functionality. Assessment of emergency equipment status occurs during installation, ensuring compliance with regulations, and during daily maintenance to prevent lapses in upkeep. Checks during installation guarantee regulatory adherence and adequate emergency preparedness, aiding construction progress tracking (Cao, Kamaruzzaman, & Aziz, 2022). In daily operations, maintenance gaps can lead to issues like inaccessible fire extinguishers or obstructed hose reels, impeding swift fire control (Xu, Chan, Leong, & Borondo, 2023). Routine inspections by authorities and community managers are essential to uphold equipment functionality (Guan, Fang, & Wang, 2018; Tse, 2002). Therefore, regular monitoring and maintenance of emergency equipment are vital for operational readiness, ensuring prompt and effective responses during crises.

During construction, delivery, daily operation, and maintenance phases, emergency equipment status checks rely on manual inspection at the current stage. However, manual checks are inefficient and costly. According to Hong Kong Fire Department guidelines (Lo, 1998), emergency equipment should be inspected annually by a registered contractor. As buildings grow larger, the number of fire safety facilities increases, straining manpower and time, potentially compromising inspection quality. This results in subpar maintenance practices (Kobes, Helsloot, de Vries, & Post, 2010). Intelligent inspection technologies (Spencer, Hoskere, & Narazaki, 2019), particularly computer vision, offer a solution by enhancing automation and data collection, potentially replacing manual labor. This technology enables automatic data recording and analysis, generating detailed reports for maintenance planning and equipment management. Ultimately, it streamlines inspections, boosts equipment reliability, and enhances maintenance efficiency.

Currently, emergency equipment management research primarily focuses on platform construction. For instance, Wang et al. (S.-H. Wang, Wang, Wang, & Shih, 2015) introduced a maintenance module for emergency equipment based on BIM. This module uses BIM's data storage to help maintenance personnel access fire safety equipment data swiftly. However, this data is manually verified, leading to concerns about accuracy. Vijayalakshmi et al. (Vijayalakshmi & Muruganand, 2017) applied IoT to monitor emergency facilities, enhancing management in two stages: improving firefighting product quality and employing Radio Frequency Identification (RFID) for equipment tracking and deficiency identification. Li et al. (Li, Becerik-Gerber, Krishnamachari, & Soibelman, 2014) proposed an intelligent emergency response framework using metaheuristic algorithms for optimal solutions, integrating emergency equipment positions for decision-making. Existing research relies heavily on manual methods, leading to outdated information on equipment status, a major challenge in equipment management. On-site emergency facility management is still emerging, facing issues due to long life cycles and high costs. Neglected concerns include facility integrity, configuration, and hazards. Damaged or outdated firefighting facilities due to neglect pose risks during emergencies. Some buildings may lack adequate or properly placed fire protection equipment, impacting safety standards. Hidden dangers like obstructed fire extinguishers pose risks if not addressed promptly. The gap between emergency facility management platforms and actual equipment status highlights the need for intelligent frameworks for equipment status detection to enhance emergency response efficiency and reliability. Addressing these issues is crucial for ensuring the effectiveness and safety of emergency responses.

While computer vision technologies have found application in civil engineering inspections, they currently face limitations in effectively assessing the status of emergency equipment. Key challenges include the scattered placement of emergency devices throughout buildings, leading to difficulties in spatially locating targets during long-distance inspections. Additionally, dynamic perspectives in video detection hinder the accurate identification and counting of objects. Moreover, the varying sizes of emergency equipment, coupled with image quality issues in videos featuring different-sized targets, necessitate detectors with enhanced performance to address these complexities.

2. METHOD

This study presents an efficient automated inspection framework that utilizes cutting-edge computer vision and deep learning technologies for accurate positioning and identification of emergency equipment. The framework not only quickly detects the status of equipment but also compiles statistics on the equipment, facilitating enhanced management and maintenance tasks.

The proposed visual-based method, depicted in Figure 1, comprises three main components: an advanced object detection network, object ID allocation and tracking, and 3D mapping with detailed device information. Initially, the framework captures both RGB and depth images using a depth camera, which are used to create an inspection map by matching 2D and 3D features. The RGB images are processed through an enhanced detection network to identify emergency equipment, while depth images are converted into a point cloud using camera parameters to establish spatial relationships. Features within each equipment's bounding box are analyzed to assign unique IDs, enabling effective tracking. These devices are then mapped onto the inspection trajectory to generate a comprehensive device distribution map. This method facilitates precise detection and tracking of emergency equipment, essential for monitoring progress, managing emergencies, and performing maintenance tasks.

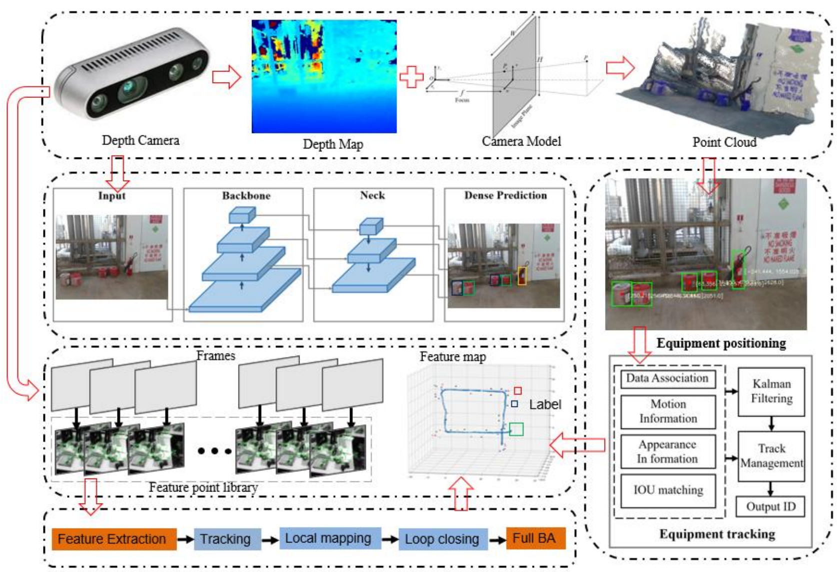


Figure 1. Fire safety equipment inspection framework

2.1 Improved Fire Safety Equipment Detection Network

Deep learning-based object detection algorithms like YOLOv5 are prevalent in industrial inspection but face challenges in detecting fire safety equipment due to issues like varying object sizes and image blurring from camera motion. To enhance performance, we propose an improved detection network that integrates C2F modules for advanced feature extraction, CPCA modules to accentuate and merge features, and DyHead to boost feature perception capabilities. This combination significantly improves detection accuracy and stability for device tracking in dynamic environments.

In the proposed network, the C2F module (Reis, Kupec, Hong, & Daoudi, 2023) is introduced instead of the C3 module. It combines three interlaced layers with 1×1 kernel and includes the DarkNet bottleneck. This module adds more skip connections, eliminates branch convolution operations, and adds splitting operations. This enriches feature information while reducing computational complexity.

The C2F module optimizes the feature extraction network's information flow and ensures lightness with gradient diversion connections. It also draws on the ELAN module's ideas to optimize the network structure for easier training. By using feature vector diversion and multi-level nested convolution, it learns multi-scale features, expands the receptive field range, and improves the network's trainability and object detection performance. The attention mechanism highlights key features, removes background interference, and fuses features effectively. The proposed network uses the CPCA attention mechanism (Huang, Chen, Zou, Lu, & Chen, 2023) to address the network's limitations in handling different-scale, -shape, and -direction information. The CPCA attention mechanism consists of two sub-modules: the channel attention module and the spatial attention module. Compared to traditional ones, CPCA has innovative designs for better capturing and enhancing key information. The proposed network adds the CPCA attention mechanism before outputting three-sized feature maps on the backbone to enhance neck section feature fusion. The model reassigns weights to different-resolution feature maps, enhancing useful features and suppressing irrelevant ones. This makes the model focus on potential fire safety equipment areas. Adding CPCA before outputting feature maps in the backbone network can globally select and weight feature map channels and spatial positions. Thus, the model can interpret the entire image's contextual information, distinguish objects and backgrounds, improving detection accuracy. The backbone network's downsampling causes object information loss, especially for small objects. Video factors also degrade object quality. To enhance head module's feature perception, DyHead (Dai et al., 2021) with multiple self-attention mechanisms is adopted. After inputting the feature map into DyHead, it becomes a 3D tensor. Features are input to scale, spatial, and task perception attention modules. DyHead unifies these perceptions, enabling the network to focus on useful info. Its cascade attention mechanism allows handling multiple tasks. Three-scale feature maps are input to a unified branch and processed by DyHead.

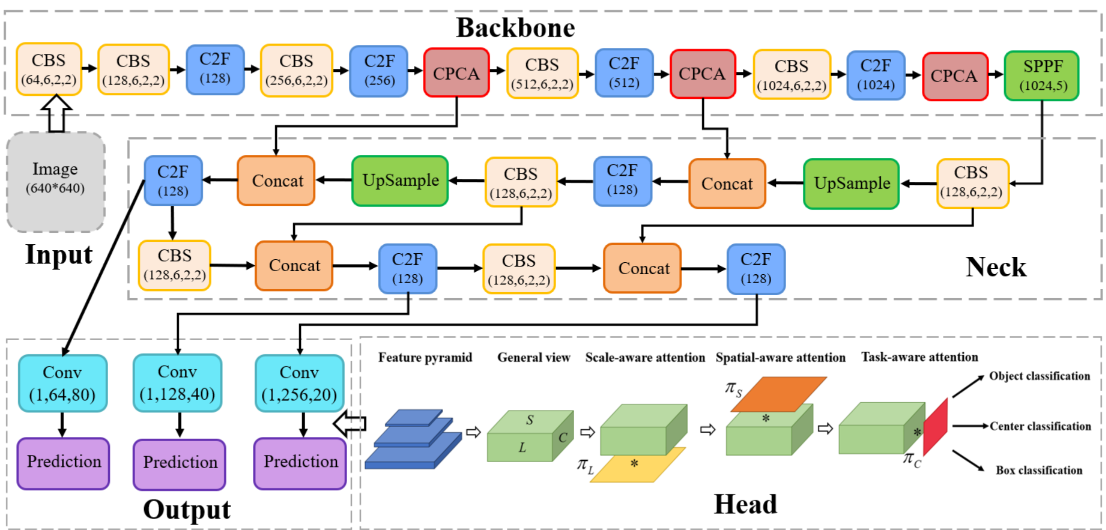


Figure 2. Enhanced architecture of the detection network

Object detection is a supervised learning task. The loss function measures the difference between predicted results and ground truth labels to guide model training. Minimizing the loss function helps object detection models learn more accurate bounding boxes and classification predictions, improving detection accuracy. For localization loss, IoU is commonly used to measure the overlap between predicted and actual bounding boxes. However, when IoU is used as a loss function, there are two issues: inability to distinguish bounding box shape differences and the problem of non-overlapping objects. For aspect ratio bounding boxes, they may have the same overlap but can't accurately surround objects. IoU can't distinguish them, resulting in the same value and inability to make reasonable weight adjustments. When predicting small objects, there may be no overlap between the ground truth and prediction box, with an IoU of 0 and a gradient of 0, making the model unable to optimize effectively.

The introduction of GIoU (Rezatofighi et al., 2019) helps alleviate the gradient problem caused by variations in the shapes of bounding boxes. The concept of minimum closure region C is added on the basis of IoU. The minimum closure region is a minimum box C that can completely enclose the two bounding boxes A and B. By incorporating the area of the minimum closure region into the calculation, GIoU can better consider the position and shape differences between bounding boxes. In addition, to address the issue of small objects, the Normalized Wasserstein Distance (NWD) (Liu, Johns, & Davison, 2019; J. Wang, Xu, Yang, & Yu, 2021) is used instead of the degree of overlap to evaluate the prediction accuracy. For most objects, their bounding boxes are usually not strictly rectangular and may contain some background pixels. In these bounding boxes, the object is often concentrated at the center of the bounding box, while the background pixels are distributed at the boundaries. To better describe the weights of different pixels in the bounding box, the bounding box is regarded as a two-dimensional Gaussian distribution. In this model, the importance of pixels from the center of the bounding box to the boundary gradually decreases to distinguish the importance of the object and the background.

The loss of the proposed $Loss_{GN}$ function can be expressed as Eq.(1). Among them, λ_G and λ_N are dynamic parameters based on the size of gradients for different loss weights. A faster decrease in loss indicates a quicker learning speed. Nevertheless, for losses with higher learning speeds, the weight values should be controlled to strike a balance among various losses (He, Zhu, Wang, Savvides, & Zhang, 2019; Liu et al., 2019; Rezatofighi et al., 2019; J. Wang et al., 2021). The dynamic learning weight is defined as Eq. (2), which represents the loss ratio at time $t-1$ and time $t-2$. Next, Equation (3) is employed to normalize the learning speed through exponential processing to obtain the weights for each type of loss. Here, B is the equilibrium coefficient that serves to adjust the degree of difference in different loss weights. As the value of B rises, the disparity in weights is reduced, leading to a more even distribution of weights. A larger B implies a greater emphasis on balancing the weights among different components or tasks.

$$Loss_{GN} = \lambda_G Loss_G + \lambda_N Loss_N \tag{1}$$

$$\lambda_k(t) = \frac{\exp(w_k(t-1)/B)}{\exp(w_G(t-1)/B) + \exp(w_N(t-1)/B)} \tag{2}$$

$$w_k(t-1) = \frac{Loss_k(t-1)}{Loss_k(t-2)}, (k = N, G) \tag{3}$$

2.2 Equipment Spatial Positioning Based on vSLAM and Improved Visual Tracking

The DeepSort algorithm (Lu et al., 2021) is used to distinguish different fire safety equipment by tracking various objects in the video and assigning unique IDs to each object. The detector provides DeepSort with information such as location, size, and appearance features. DeepSort uses the kalman filtering algorithm to predict the position and state of objects in the current frame based on the parameters detected in the previous frame. It also uses deep neural networks to extract object RGB features. Cosine distance measures the similarity of appearance features, while mahalanobis distance calculates the similarity of motion features to form a cost matrix. Hungarian algorithm evaluation cost matrix. If the matching distance is less than the predefined threshold, it is considered that the IDs are the same, indicating a successful match. For mismatched trajectories, perform secondary matching based on IoU. If successful, use the kalman filter to update the object state. Determine whether to remove based on the trajectory that does not match the lifespan assessment. Unlike previous tracking tasks that used mobile cameras as inspection sensors, this results in more intense object motion in the image. The IoU matching strategy based on bounding box position similarity is prone to errors or omissions. Additional object depth information can improve tracking stability. Considering the uniform variation of distance between objects and cameras, we introduce a depth change rate based on depth cameras to optimize the data association evaluation method. When the matching distance is near the threshold, depth information can be used as an auxiliary judgment to improve the robustness of tracking accuracy.

Once the 2D bounding box and ID details of every fire safety equipment are obtained, the next step is to get the 3D position information of the equipment with respect to the camera. Since a depth camera is used, the depth value d of each pixel in the image is known. The camera model allows us to combine the camera's internal parameters and depth values to obtain the 3D coordinates of any pixel point in the camera coordinate system. Given that the framework only needs an approximate position of the fire safety device, which is sufficient for subsequent maintenance or information updates on the management platform. Hence, we take the center of the 2D bounding box as input to acquire the 3D coordinates of that point, representing the 3D positional information of the equipment. The projection process is presented. For the single frame, only the spatial transformation relationship (R_{wc}, T_{wc}) from the camera coordinate system $O_c - xyz$ to the 3D coordinate system $O_w - xyz$ needs to be obtained, and the target point can be projected into world coordinates by rigid transformation.

Fire safety equipment in buildings is scattered in distribution. As a result, cameras need to obtain the status of all such equipment from multiple frames. Hence, it is necessary to establish a unified world coordinate system and convert the fire safety equipment detected from each perspective into this coordinate system. In the case of multiple frames, when dealing with multiple frames, it is crucial to determine the spatial transformation relationship $(R_{k_i k_{i+1}}, T_{k_i k_{i+1}})$ between adjacent frames, which indicates the camera's spatial pose at each moment. In order to obtain the set of spatial poses of the camera at different timestamps, the proposed framework computes the correspondence between feature points in adjacent images to determine the camera's pose during motion, which includes transformation and rotation. Suppose the rigid spatial variation matrix between the two frames of the camera at time k and time $k+1$ is $T_{k:k-1} \in \mathbb{R}^{4\times4}$, it can be expressed follows:

$$T_{k:k-1} = \begin{bmatrix} R_{k:k-1} & t_{k:k-1} \\ 0 & 1 \end{bmatrix} \tag{4}$$

where, $R_{k:k-1} \in \mathbb{R}^{3\times3}$ is the rotation matrix and $t_{k:k-1} \in \mathbb{R}^{3\times1}$ is the translation matrix. As can be seen in Figure 5(c), along the inspection path of the camera. By setting $T_{1:k} = \{T_{1:0}, T_{2:1}, \cdots, T_{k:k-1}\}$ as the set of cameras pose transformations, it can be achieved through spatial coordinate transformation. It is assumed that at a certain time step k, the spatial coordinate of any point under the camera's perspective can be converted into a unified world coordinate system by using Equation (5), as shown below.

$$\begin{bmatrix} x_{pw} \\ y_{pw} \\ z_{pw} \\ 1 \end{bmatrix} = T_{1:0} \cdot T_{2:1} \cdots T_{k-1:k-2} \cdot T_{k:k-1} \begin{bmatrix} x_p^k \\ y_p^k \\ z_p^k \\ 1 \end{bmatrix} \tag{5}$$

To attain higher accuracy and robustness, our framework employs one of the most advanced visual SLAM algorithms, ORB-SLAM3 (Campos, Elvira, Rodriguez, M. Montiel, & D. Tardos, 2021) to obtain $(R_{k_i k_{i+1}}, T_{k_i k_{i+1}})$. Moreover, for the purpose of obtaining a real-scaled camera trajectory, the RGB and depth maps gathered by the depth camera are employed as inputs to the SLAM system. Within this system, first, all 3D coordinates related to the same device ID are retrieved. Then, after setting a threshold to remove outliers, the average of the remaining points is regarded as the 3D position of the equipment. Eventually, by projecting the detected devices onto the SLAM map, a map containing 3D device location information can be constructed, as depicted in Figure 3.

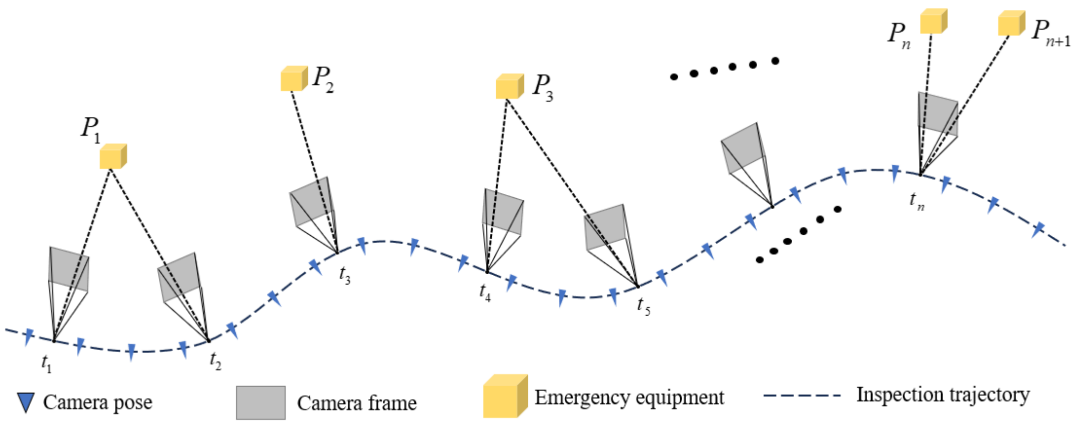


Figure 3. Creation of an equipment information map

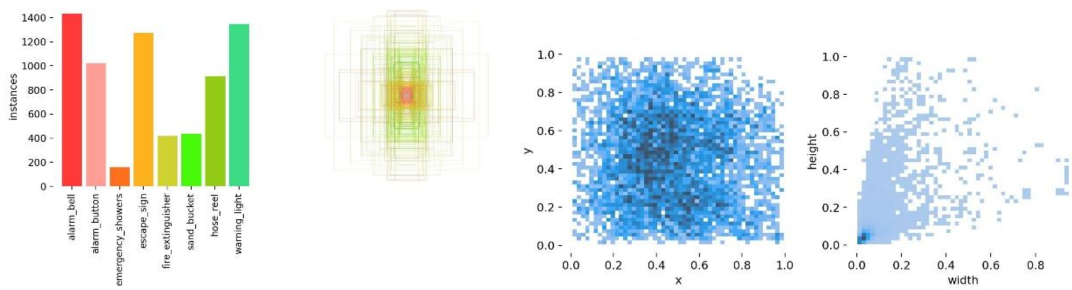
3. RESULTS

3.1 Dataset and training

The existing public datasets are insufficient for meeting the training and testing requirements of this study as they lack consideration of an adequate amount of emergency equipment. The study identified eight crucial emergency equipment types: alarm bell, alarm button, emergency shower, escape sign, fire extinguisher, sand bucket, hose reel, and warning light. To address this gap, a dataset featuring 2793 instances of diverse emergency equipment was meticulously curated. Images, captured using smartphones or D435i depth cameras under varying lighting conditions, were intentionally marred by occlusion, lighting inconsistencies, and motion blur to simulate real-world challenges. Leveraging the LabelImg annotation tool, bounding boxes were meticulously marked around each item, followed by data augmentation using Mosaic to enhance dataset diversity. Figure 4 presents several annotated and enhanced images. The dataset was then partitioned randomly into training, validation, and testing subsets, each comprising 70%, 15%, and 15% of the total images, respectively.

Figure 4. The dataset after data augmentation

In Figure 5, information regarding bounding boxes in the dataset is depicted. Figure 5(a) illustrates the object quantities for each type within the dataset, primarily influenced by the frequency of emergency equipment use and their placements in buildings. Meanwhile, Figure 5(b) showcases the diverse sizes and aspect ratios of object bounding boxes. The distribution of the center points of the bounding boxes is depicted in Figure 5(c), highlighting their concentration in the central region of the image data. Finally, Figure 5(d) displays a scatter plot representing the width and height of the bounding boxes, with the darkest hues clustered in the bottom left corner, indicating a prevalence of small targets within the dataset.



(a) (b) (c) (d)

Figure 5. Information about the manually labeling of objects in dataset.

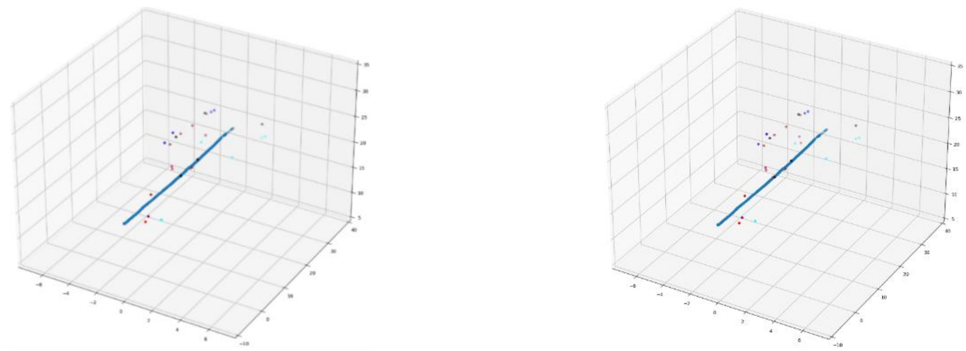
3.1 Experimental validation

To assess the effectiveness of the proposed method and establish a direct link between equipment status and the corresponding devices, four distinct corridor scenarios were selected for on-site experiments (refer to Figure 6). Corridors, typically housing a high density of fire safety equipment, were chosen due to their significance in this context. Scenario ① features a straight corridor with consistent artificial lighting, offering a uniform background that facilitates the easy identification of fire safety equipment in a structured environment. In contrast, scenario ②, also a straight corridor, transitions from an outdoor section with strong natural light to an indoor area with relatively weaker illumination, potentially causing fire safety equipment to be obscured by debris. Finally, scenario ③ presents an indoor Z-shaped corridor with even artificial and natural lighting. Despite this, the corridor's spaciousness results in some emergency maps appearing as small objects in the images.

(a) Scenario ① (b) Scenario ② (c)Scenario ③

Figure 6. Four on-site scenarios

Figures 7 to 9 show the distribution of camera inspection paths and fire safety equipment, comparing the results of the original method and the improved method. The statistical results are shown in Tables 1-3, respectively. Both the original method and the improved method can effectively cope with this scenario ①, counting the correct number, thanks to the strong contrast backgrounds. In scenario ②, due to the issue of obstruction and debris, the original method is unable to detect one of the fire extinguishers, while poor lighting caused an alarm bell to be missed. The improved method can detect these two ignored devices. In scenario ③, the improved method can also provide similar improvements, especially for smaller emergency devices such as alarm bells and warning lights.

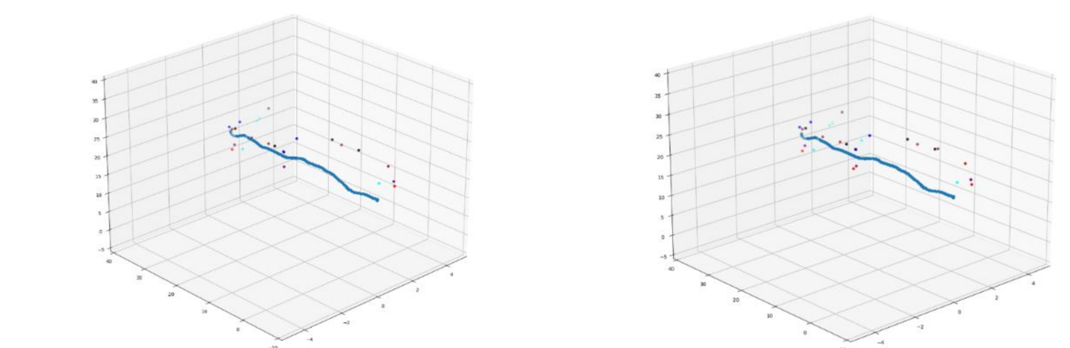


(a) Inspection map of the original method (b) Inspection map of the improved method

Figure 7 Inspection map for scenario①

Table 1. Scenario ① fire safety equipment inspection statistics

Scenario ①	Ground truth	Original method	Improved method
Alarm_bell	3	3	3
Alarm_button	2	2	2
Emergency_showers	0	0	0
Escape_sign	7	7	7
Fire_extinguisher	0	0	0
Sand_bucket	0	0	0
Hose_reel	2	2	2
Warning_light	0	0	0

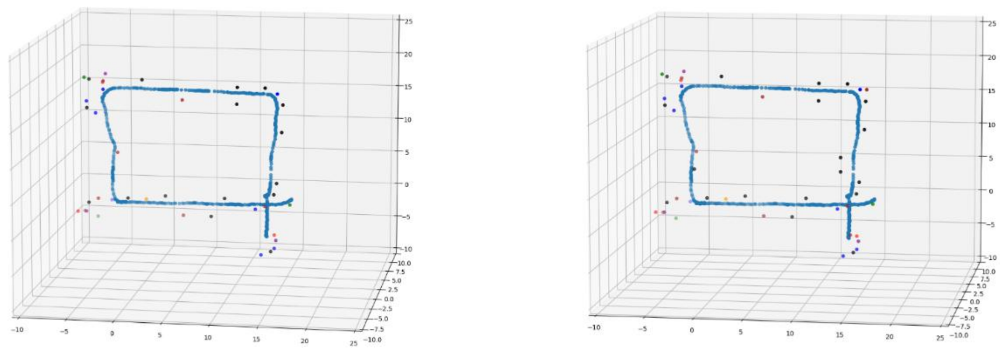


(a) Inspection map of the original method (b) Inspection map of the improved method

Figure 8 Inspection map for scenario②

Table 2. Scenario ② fire safety equipment inspection statistics

Scenario ②	Ground truth	Original method	Improved method
Alarm_bell	***7***	*6*	***7***
Alarm_button	3	2	3
Emergency_showers	0	0	0
Escape_sign	4	4	4
Fire_extinguisher	***5***	*4*	***5***
Sand_bucket	0	0	0
Hose_reel	3	3	3
Warning_light	5	5	5



(a) Inspection map of the original method (b) Inspection map of the improved method

Figure 9 Inspection map for scenario ③

Table 3. Scenario ③ fire safety equipment inspection statistics

Scenario ④	Ground truth	Original method	Improved method
Alarm_bell	***9***	*7*	***9***
Alarm_button	3	3	3
Emergency_showers	3	3	3
Escape_sign	8	8	8
Fire_extinguisher	0	0	0
Sand_bucket	1	1	1
Hose_reel	3	3	3
Warning_light	***19***	*16*	***19***

4. DISCUSSION

Buildings often face the issue of fire safety equipment malfunctioning due to insufficient maintenance (Dong et al., 2014). As building scales increase, the workload for regular inspections grows significantly, necessitating substantial manpower and time. To address this, it is crucial to invest in automated inspection and maintenance systems for fire safety equipment to ensure they remain in optimal working condition, providing reliable safety for the building and its occupants. We have developed an advanced detection framework that leverages vSLAM to construct a three-dimensional relationship among equipment, cameras, and trajectories, effectively pinpointing the spatial positions of dispersed devices. DeepSort is utilized to track objects in video feeds, assigning unique IDs and enabling accurate object counting and identity recognition. Additionally, feature enhancement strategies and improvements to the loss function have been integrated to boost the detection network's performance, minimizing missed detections and false positives. The results demonstrate that this method adeptly handles various inspection scenarios, significantly improving accuracy and reliability of the system.

5. CONCLUSIONS

Compared to the original method, the improved method significantly enhances the robustness and accuracy of emergency equipment detection, offering more reliable support for emergency management and response. Looking ahead, developing an integrated management platform for fire safety equipment could be a priority. This platform would integrate with existing detection frameworks to provide comprehensive oversight and operational support for equipment. It could include features like intelligent scheduling and advanced path planning algorithms to streamline inspection processes, thus reducing inspection times and distances traveled. Furthermore, the integration of mobile terminals and apps could transform inspection management, enabling robots to remotely handle and execute tasks. This would allow inspection personnel to seamlessly receive task allocations, access optimized path plans, retrieve device-specific information, and easily record and submit inspection results. Such a holistic approach promotes real-time data sharing, boosts work efficiency and precision, and ultimately improves operational workflows in fire safety equipment management.

REFERENCES

Campos, C., Elvira, R., Rodriguez, J. J. G., M. Montiel, J. M., & D. Tardos, J. (2021). ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Transactions on Robotics, 37*(6), 1874-1890. doi:10.1109/tro.2021.3075644

Cao, Y., Kamaruzzaman, S. N., & Aziz, N. M. (2022). Building Information Modeling (BIM) Capabilities in the Operation and Maintenance Phase of Green Buildings: A Systematic Review. *Buildings, 12*(6). doi:10.3390/buildings12060830

Dai, X., Chen, Y., Xiao, B., Chen, D., Liu, M., Yuan, L., & Zhang, L. (2021). *Dynamic head: Unifying object detection heads with attentions.* Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Dong, Q., You, F., & Hu, S.-q. (2014). Investigation of Fire Protection Status for Nanjing Representative Historical Buildings and Future Management Measures. *Procedia Engineering, 71*, 377-384. doi:10.1016/j.proeng.2014.04.054

Guan, Y.-x., Fang, Z., & Wang, T.-r. J. P. e. (2018). Fire risk assessment and daily maintenance management of cultural relic buildings based on ZigBee technology. *Procedia Engineering, 211*, 192-198. doi:10.1016/j.proeng.2017.12.004

He, Y., Zhu, C., Wang, J., Savvides, M., & Zhang, X. (2019). *Bounding box regression with uncertainty for accurate object detection.* Paper presented at the Proceedings of the ieee/cvf conference on computer vision and pattern recognition.

Huang, H., Chen, Z., Zou, Y., Lu, M., & Chen, C. J. a. p. a. (2023). Channel prior convolutional attention for medical image segmentation. *Computers in Biology and Medicine*. doi:10.1016/j.compbiomed.2024.108784

Ivanov, M. L., & Chow, W.-K. (2022). Fire safety in modern indoor and built environment. *Indoor and Built Environment, 32*(1), 3-8. doi:10.1177/1420326x221134765

Kobes, M., Helsloot, I., de Vries, B., & Post, J. G. (2010). Building safety and human behaviour in fire: A literature review. *Fire Safety Journal, 45*(1), 1-11. doi:10.1016/j.firesaf.2009.08.005

Li, N., Becerik-Gerber, B., Krishnamachari, B., & Soibelman, L. J. A. i. C. (2014). A BIM centered indoor localization algorithm to support building fire emergency response operations. *Automation in Construction, 42*, 78-89. doi:10.1016/j.autcon.2014.02.019

Liu, S., Johns, E., & Davison, A. J. (2019). *End-to-end multi-task learning with attention.* Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Lo, S. M. J. S. S. (1998). A building safety inspection system for fire safety issues in existing buildings. *Structural Survey, 16*(4), 209-217. doi:10.1108/02630809810243220

Lu, T., Tervola, S., Lü, X., Kibert, C. J., Zhang, Q., Li, T., & Yao, Z. (2021). A novel methodology for the path alignment of visual SLAM in indoor construction inspection. *Automation in Construction, 127*. doi:10.1016/j.autcon.2021.103723

Pan, Y., Braun, A., Brilakis, I., & Borrmann, A. (2022). Enriching geometric digital twins of buildings with small objects by fusing laser scanning and AI-based image recognition. *Automation in Construction, 140*. doi:10.1016/j.autcon.2022.104375

Reis, D., Kupec, J., Hong, J., & Daoudi, A. J. a. p. a. (2023). Real-Time Flying Object Detection with YOLOv8. *arXiv preprint, arXiv:2305.09972*. doi:10.48550/arXiv.2305.09972

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). *Generalized intersection over union: A metric and a loss for bounding box regression.* Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Spencer, B. F., Hoskere, V., & Narazaki, Y. (2019). Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering, 5*(2), 199-222. doi:10.1016/j.eng.2018.11.030

Tang, R., Fitzgerald, G., Hou, X. Y., & Wu, Y. P. (2014). Building an evaluation instrument for China's hospital emergency preparedness: a systematic review of preparedness instruments. *Disaster Med Public Health Prep, 8*(1), 101-109. doi:10.1017/dmp.2014.10

Tse, P. W. J. J. o. Q. i. M. E. (2002). Maintenance practices in Hong Kong and the use of the intelligent scheduler. *Journal of Quality in Maintenance Engineering, 8*(4), 369-380. doi:10.1108/13552510210448540

Vijayalakshmi, S., & Muruganand, S. J. I. R. J. E. T. (2017). Internet of Things technology for fire monitoring system. *4*(6), 2140-2147.

Walters, M., & Hastings, E. J. F. (1998). Fire safety legislation in Hong Kong. *Facilities, 16*(9/10), 246-253. doi:10.1108/02632779810229066

Wang, D., Gao, L., Zheng, J., Xi, J., & Zhong, J. (2025). Automated recognition and rebar dimensional assessment of prefabricated bridge components from low-cost 3D laser scanner. *Measurement, 242*. doi:10.1016/j.measurement.2024.115765

Wang, J., Xu, C., Yang, W., & Yu, L. J. a. p. a. (2021). A normalized Gaussian Wasserstein distance for tiny object detection. *arXiv preprint, arXiv:2110.13389*. doi:10.48550/arXiv.2110.13389

Wang, S.-H., Wang, W.-C., Wang, K.-C., & Shih, S.-Y. (2015). Applying building information modeling to support fire safety management. *Automation in Construction, 59*, 158-167. doi:10.1016/j.autcon.2015.02.001

Xin, J., & Huang, C. (2013). Fire risk analysis of residential buildings based on scenario clusters and its application in fire risk management. *Fire Safety Journal, 62*, 72-78. doi:10.1016/j.firesaf.2013.09.022

Xu, W., Chan, S. C., Leong, W. Y., & Borondo, F. (2023). Effectiveness Study of Artificial Intelligent Facility System in Maintaining Building Fire Safety (Case Study: Typical Public Building Cases of Fire-Fighting Facilities Management in China). *Discrete Dynamics in Nature and Society, 2023*, 1-21. doi:10.1155/2023/2592322