The Indonesian Journal of Computer Science
www.ijcs.net
Volume 13, Issue 6, December 2024
https://doi.org/10.33022/ijcs.v13i6.4541

Fire and Smoke Detection with 3D Position Estimation Using YOLO

Bryan Anthony*[1], Richard Evan Sutanto[2], Yuwono Marta Dinata[3]
bryan.wansen@gmail.com*[1], richardwenz91@gmail.com[2], yuwono.dinata@ciputra.ac.id[3]
[1,2,3] Informatics, Universitas Ciputra, Surabaya, Indonesia

Article Information

Received : 11 Dec 2024
Revised : 15 Dec 2024
Accepted : 30 Dec 2024

Keywords

Computer Vision, Fire And Smoke, Fire Detection, Machine Learning, 3D Localization

Abstract

Fire and smoke detection systems are essential for safety, but traditional methods often face issues like false alarms and poor localization accuracy. This study integrates advanced object detection models, a confidence-based voting mechanism, and 3D localization to address these challenges. Using three cameras, the system detects fire or smoke and estimates its 3D position (x, y, z) through bounding box depth estimation and camera placement. A voting mechanism enhances reliability by requiring validation from at least two cameras with a confidence threshold of 0.5. YOLOv5s achieved 92% accuracy, 96% precision, 95% recall, mAP50 of 98%, and 87.02 FPS, making it suitable for real-time use. YOLOFM-NADH+C3 offered comparable accuracy (92%) but better localization precision with a 0.22 cm error versus YOLOv5s' 1.53 cm, albeit at a slower FPS (54.82). Experiments confirm the system's ability to reduce false positives and localize fire/smoke accurately under challenging conditions.

A. Introduction

A fire is an event that occurs when uncontrolled flames and heat damage objects or the surrounding environment. Fires pose severe risks, leading to substantial property damage and endangering lives [1]. Detecting flames and smoke in real-time is essential to minimize these risks and enable prompt emergency responses. Smoke inhalation is a leading cause of death in fire incidents, underscoring the importance of early detection systems.

Traditional fire detection systems have relied heavily on sensor technologies, such as multi-sensor fusion and infrared sensors. However, these methods are often limited by geographical coverage and high implementation costs [2]. Conventional smoke detectors also have notable limitations, including a tendency to trigger false alarms, which can disrupt businesses and reduce productivity. According to the UK government's report, fire and rescue services responded to 594,384 incidents in the year ending June 2024, of which only 22% were actual fires, while a significant 43% were false alarms. This consumes valuable resources and imposes unnecessary costs on both fire services and businesses [3]. These challenges highlight the need for intelligent fire alarm systems to reduce false alarms and improve detection efficiency. Despite a general decrease in fire incidents, the high frequency of false alarms underscores the limitations of traditional systems [4].

Recent advancements in computer vision, especially deep learning, have shown promise in addressing these challenges by offering accurate and reliable detection capabilities. By leveraging artificial intelligence and machine learning, these systems analyze visual data in real-time to quickly identify signs of smoke and fire, significantly enhancing safety measures [5]. Currently, sensor-based systems, image processing, and machine learning are employed in IoT monitoring systems to detect fire and smoke. Computer vision-based systems, in particular, have demonstrated promising results and garnered attention for their precision and reliability in detection [6].

Fire detection using deep learning and computer vision has been extensively studied, with various models achieving significant success in detecting fire and smoke under diverse conditions. Among these, the YOLO series (You Only Look Once) has emerged as a prominent choice for real-time object detection due to its balance of speed and accuracy. For example, Muhammad et al. leveraged YOLOv8 and YOLOv7 to achieve a mean average precision (mAP@50) of 92.6%, precision of 83.7%, and recall of 95.2%, highlighting their effectiveness in fire detection tasks [7]. However, challenges related to real-time processing and computational constraints persist [8].

Similarly, older YOLO versions, such as YOLOv2, demonstrated high accuracy in specific applications, achieving 96.82% accuracy in fire and smoke detection for anti-fire surveillance systems [9]. Later versions, such as YOLOv5, introduced architectural improvements that enhanced performance for detecting small targets and operating in dynamic environments [10]. This balance of precision and speed makes YOLOv5 particularly suitable for real-time fire detection systems in resource-constrained environments [11].

Advancements in YOLO models, such as YOLOFM, have addressed specific limitations like small-target detection and fire-smoke differentiation by incorporating features like FocalNext, quantization-aware processing (QAHARep-

FPN), and NADH decoupled heads [12]. These modifications make YOLOFM highly effective in distinguishing fire from smoke, especially in complex indoor settings where precise localization is critical. Multi-camera systems further enhance fire detection by improving spatial positioning and enabling 3D localization. These systems reduce occlusions and provide accurate object tracking across viewpoints, as demonstrated by Amosa et al. [13] and Kadam et al. [14]. Advanced frameworks like Conditional Random Field (CRF) for 3D position estimation [15] and tracking approaches such as DeepLabCut [16] underscore the significance of spatial accuracy in fire detection.

Despite recent progress in fire detection, existing models often struggle with precise spatial localization and real-time processing in resource-constrained environments. This study bridges these gaps by integrating YOLOv5s and YOLOFM with a multi-camera system for 3D position estimation and a voting mechanism to reduce false alarms. YOLOv5 provides lightweight, real-time detection capabilities, while YOLOFM excels in recognizing small targets and differentiating fire from smoke. The proposed system addresses key challenges like occlusions and detection precision in complex environments, contributing to faster and more accurate emergency responses.

B. Research Method

The flow of the research is depicted in Figure 1.

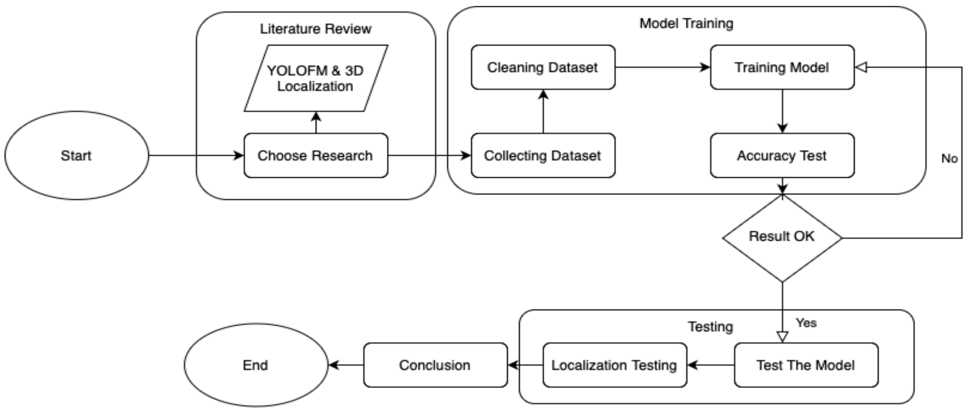


Figure 1. Research Flow

System Architecture

Our proposed system, illustrated in Figure 2, addresses the challenges of false alarms and enhances early awareness of fire and smoke incidents through a multi-camera approach and advanced processing techniques. The system ensures reliable and precise monitoring in real-world scenarios by integrating object detection, validation, and 3D localization.

1. Object Detection with Multiple Cameras

The first step involves detecting fire or smoke using YOLO-based object detection models on images of multiple cameras strategically positioned at different angles. Each camera independently processes its frames using the same YOLO model to identify fire or smoke within 2D bounding boxes in its respective field of view.

This uniform application of the detection model ensures consistency in identifying potential fire or smoke events across all cameras.

2. Voting Mechanism for Validation

Each camera's detected fire or smoke positions are passed to a voting mechanism. This module cross-references detections across multiple cameras to confirm the presence of fire or smoke, using a confidence threshold of 0.5 and ensuring consistency among detections. By requiring at least two cameras to validate a detection, the voting mechanism significantly reduces the likelihood of false positives, allowing only reliable events to proceed to the next analysis stage.

3. 3D Localization for Precise Positioning

The validated bounding boxes and position estimates are processed to compute the detected fire or smoke's precise 3D coordinates (x, y, z). This spatial localization is achieved using geometric transformations derived from each camera's configuration, such as field of view, tilt angle, and position in the room. The calculated 3D coordinates accurately represent the object's location, enabling timely and targeted responses. Accurate 3D localization provides critical information for activating alarms, notifying response teams, or efficiently managing resources.

The system effectively addresses the challenges of false alarms and limited spatial accuracy by combining robust detection, validation, and localization mechanisms. It ensures precise detection, reliable validation, and accurate positioning, offering a practical and scalable fire and smoke monitoring system solution.

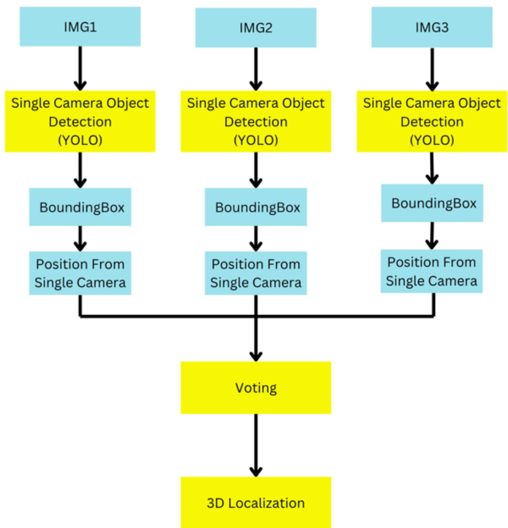


Figure 2. Fire Detection and Localization Workflow Using YOLO and 3D Position Estimation

Dataset and Preprocessing

The dataset utilized in this study is derived from the FM-VOC18644 dataset, a collection specifically designed to enhance the detection of fire and smoke objects using deep learning models. Although the original FM-VOC18644 dataset consist of 18,644 images, only 16,844 were successfully downloaded. Further cleaning was conducted to ensure the data's usability and quality, removing corrupted or unsuitable images. After processing, the dataset was divided into three subsets following an 80:10:10 split: 12,712 images (80%) for training, 1,410 images (10%) for validation, and 1,567 images (10%) for testing. This division ensures a balanced approach for model development and evaluation, enabling robust performance assessment across various scenarios. An example of our dataset images can be seen in Figure 3.

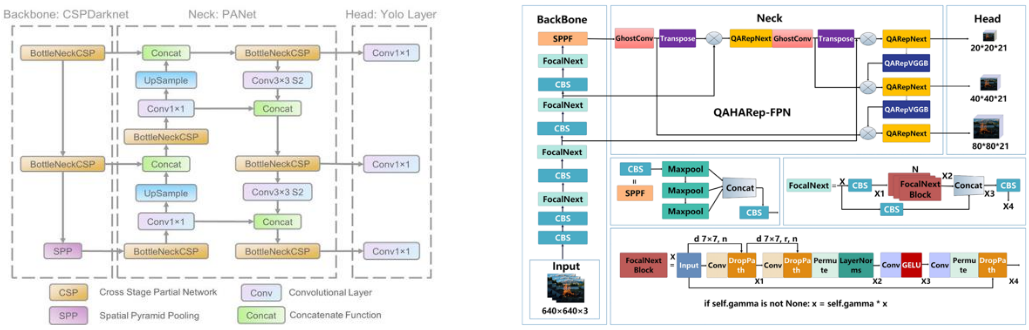
Figure 3. Image Examples From The FM-VOC18644 Dataset With And Without Labels

This study implemented preprocessing measures as described in the referenced journal, which include data augmentation techniques such as image flipping, rotation, and brightness adjustments to increase data quality and variety. Following these initial steps, a further data cleansing phase was applied to ensure the dataset's applicability for model training. This included validating image integrity and confirming size consistency, aiming to optimize data usability for accurate training results.

Model Architecutre

In this experiment, we use YOLOv5 and YOLOFM for fire and smoke detection due to their balance of accuracy and computational efficiency, making them suitable for real-time applications. YOLOv5 employs a CSP backbone for efficient feature extraction, a PANet neck for multi-scale feature fusion, and a detection head for precise localization and classification across small, medium, and large objects. Techniques like mosaic augmentation and adaptive anchor generation further enhance its real-time performance.

YOLOFM builds on YOLOv5n with enhancements tailored for fire and smoke detection, such as the FocalNext backbone for multi-scale feature extraction, the QAHARep-FPN neck for efficient handling of complex environments, and the NADH decoupled head for improved task separation and faster convergence. It uses Focal-SIoU loss for refined bounding box regression, achieving higher precision and recall compared to YOLOv5n. Figures 4 show the architectures of YOLOv5 and YOLOFM.



(a) (b)

Figure 4. (a) YOLOv5 Architecture [17] (b) YOLOFM Architecture [12]

Voting System

The voting system ensures reliable fire or smoke detection by aggregating confidence scores (Ci) from multiple cameras. For each frame, YOLO detects fire or smoke and assigns confidence scores. A detection is validated (V = 1) if at least T cameras report score above the threshold θ; otherwise, V = 0. Here we use soft voting where we adjust the θ as 0.5.

The voting threshold T is calculated dynamically: T=⌈N/2⌉ for odd N(majority) and T = (N/2) + 1 for even N (strict majority), where N is the total number of cameras. This approach reduces false alarms by requiring multi-camera consensus. The voting can be seen in Eq. (1) and Eq. (2).

$$V = \begin{cases} 1, & if \sum_{i=1}^{N} II(Ci \geq \theta) \geq T \\ 0, otherwise \end{cases}$$
(1)

$$T = \begin{cases} \lceil N/2 \rceil, & if\ N\ is\ odd \\ (N/2) + 1, & if\ N\ is\ even \end{cases}$$
(2)

3D Localization

The 3D localization method calculates the spatial position (x, y, z) of objects using trigonometry and single-camera geometry. Each camera independently computes local coordinates using its height, tilt angle, and the detected object's bounding box coordinates. The angular displacement of the object, θ_x and θ_y, is derived using field of view (FOV) and image dimensions as shown in Eq. (3) and Eq. (4).

$$\theta_x = \left(x_{2D} - \frac{image\ width}{2}\right) \cdot \frac{FOV\ Horizontal}{image\ width}$$
(3)

$$\theta_y = \left(y_{2D} - \frac{image\ height}{2}\right) \cdot \frac{FOV\ Vertical}{image\ height}$$
(4)

The horizontal distance d is calculated using trigonometry Eq. (5), accounting for camera height and tilt angle, while the local z-coordinate is determined by Eq. (6).

$$d = \frac{camera\ height}{\tan(tilt\ angle + \theta_y)} \tag{5}$$

$$z = camera\ height + d \cdot \sin(tilt\ angle + \ \theta_y) \tag{6}$$

Local coordinates are converted to a global coordinate system anchored at Camera 1. Adjustments for Cameras 2 and 3 are made based on their known positions. A voting mechanism ensures accuracy by integrating detections from multiple cameras. If at least two cameras detect the object with confidence scores above a predefined threshold, the global 3D position is averaged using Eq. (7), Eq. (8), and Eq. (9). Otherwise, the object is marked undetected.

$$x_{final} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{7}$$

$$y_{final} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{8}$$

$$z_{final} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{9}$$

Evaluation Method

The YOLOv5-based fire and smoke detection model and the 3D localization system were evaluated using key metrics to assess detection accuracy, localization precision, and computational efficiency. Detection metrics include precision, recall, and mean average precision (mAP). Precision, calculated using Eq. (10), measures the proportion of true positives to total positive predictions. Recall, defined in Eq. (11), quantifies the model's ability to identify all true instances of fire or smoke. mAP50 evaluates the precision at an Intersection over Union (IoU) threshold of 0.5, while mAP50-95 averages precision over multiple IoU thresholds from 0.5 to 0.95, providing a comprehensive measure of detection performance. Efficiency metrics include Frames Per Second (FPS), which calculates the number of frames processed per second during inference, and overall accuracy, which combines precision and recall to assess prediction correctness (Eq. (12)). Model complexity is evaluated based on trainable parameters (M) and GFLOPs (G), representing the computational requirements per image. For localization, the localization error is calculated using Eq. (13) to quantify the difference between the ground truth location ($x_{true}, y_{true}, z_{true}$) and the estimated location ($x_{pred}, y_{pred}, z_{pred}$), derived through the voting mechanism. These metrics collectively ensure a thorough evaluation of the model's ability to balance accuracy, speed, and computational demands, making it suitable for real-world, resource-constrained environments.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$
(10)

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$
(11)

$$Overall\ Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Predictions}$$
(12)

$$Localization\ Error = \sqrt{\left(x_{true} - x_{pred}\right)^2 + \left(y_{true} - y_{pred}\right)^2 + \left(z_{true} - z_{pred}\right)^2}$$
(13)

C. Result and Discussion

This section presents the outcomes of the experiments conducted to evaluate the YOLOv5-based fire and smoke detection model and the 3D localization system. The results are analyzed based on detection performance, computational efficiency, and localization accuracy.

Model Results

To evaluate the performance of the fire and smoke detection system, a series of experiments were conducted using both original YOLOv5 models (YOLOv5n, YOLOv5s, YOLOv5m, and YOLOv5l) and their modified versions. Custom YOLOFM-based models were also tested, incorporating architectural changes aimed at enhancing detection accuracy and computational efficiency. One of the key modifications involved the YOLOFM architecture, which was tested extensively with multiple adjustments. Variants such as YOLOFM-SPPF-LLR introduced multiple SPPF layers while lowering the learning rate to stabilize training and improve performance. YOLOFM-Conv replaced GhostConv with standard convolution to analyze the trade-offs between computational cost and detection accuracy. YOLOFM-NADH+C3 further streamlined the architecture by replacing GhostConv and QARepNeXt with Conv, Upsample, and C3 modules, simplifying the head and backbone without compromising detection capabilities. Additional variants, such as YOLOFM-FocalNext and YOLOFM-Freeze, explored the integration of FocalNext for improved feature representation and freezing selected layers to fine-tune performance.

Each model was trained and evaluated under the same conditions, using identical datasets and hyperparameters. Metrics such as precision, recall, mAP50, mAP50-95, FPS, overall accuracy, parameters, and GFLOPs were used to assess the models' performance comprehensively. The results of these experiments are presented in Table 1 and Table 2.

Table 1 summarizes the performance of YOLOv5-based models, showcasing both the original architectures and modified versions that included enhancements such as NADH, FocalNext, and QAHARep-FPN integration. Table 2 focuses on YOLOFM-based models, detailing the impact of structural adjustments like replacing GhostConv with Conv and integrating NADH and C3 modules. These comparisons reveal the trade-offs between detection accuracy, computational efficiency, and overall performance.

Table 1. YOLOv5-based Model

Model	Precision	Recall	mAP50	mAP50-95	FPS	Overall Accuracy
YOLOv5nNADH	0.95	0.94	0.97	0.73	60.2	0.89
YOLOv5nFocalNext	0.93	0.93	0.97	0.68	79.97	0.87
YOLOv5nQAHARep-FPN	0.93	0.92	0.96	0.69	65.67	0.86
YOLOv5n	0.93	0.91	0.95	0.63	83.08	0.86
YOLOv5s	**0.96**	**0.95**	**0.98**	**0.80**	**87.02**	**0.92**
YOLOv5m	0.95	0.93	0.96	0.79	73.30	0.88
YOLOv5l	0.97	0.94	0.97	0.90	55.13	0.91
YOLOv5sNADH	0.93	0.92	0.96	0.70	79.02	0.87
YOLOv5sQAHARep-FPN	0.95	0.93	0.96	0.73	66.13	0.88
YOLOv5sFocalNext	0.96	0.95	0.98	0.78	82.96	0.91
YOLOv5sNADHFOCAL	0.95	0.94	0.98	0.75	77.59	0.90

Table 2. YOLOFM Variants

Model	Precision	Recall	mAP50	mAP50-95	FPS	Overall Accuracy
YOLOFM	0.94	0.94	0.98	0.73	44.02	0.89
YOLOFM-SPPF-LLR	0.92	0.90	0.95	0.66	44.89	0.84
YOLOFM-Conv	0.94	0.94	0.97	0.72	57.08	0.89
YOLOFM-NADH+C3	**0.95**	**0.96**	**0.98**	**0.75**	**54.82**	**0.92**
YOLOFM-FocalNext	0.94	0.95	0.97	0.72	63.11	0.90
YOLOFM-Freeze	0.84	0.85	0.89	0.56	45.41	0.72

After analyzing the results, YOLOv5s and YOLOFM-NADH+C3 emerged as the top-performing models in terms of overall accuracy, detection precision, and recall. YOLOv5s achieved the highest overall accuracy of 0.92, with precision of 0.96, recall of 0.95, and mAP50 of 0.98. YOLOFM-NADH+C3 demonstrated comparable accuracy (0.92) while also excelling in localization tasks. Table 3 presents the parameters and GFLOPs of each model to compare computational efficiency. These metrics provide additional insights into the trade-offs between computational cost and detection performance.

Table 3. Model Computational Comparison

Model	Parameters(M)	GFLOPs (G)
YOLOv5nNADH	3.92	6.25

Model	Parameters(M)	GFLOPs (G)
YOLOv5nFocalNext	1.76	2.07
YOLOv5nQAHARep-FPN	2.26	2.81
YOLOv5n	1.77	0.65
YOLOv5s	**7.02**	**7.88**
YOLOv5m	20.86	23.93
YOLOv5l	46.11	53.83
YOLOv5sNADH	3.63	5.78
YOLOv5sQAHARep-FPN	9.03	10.83
YOLOv5sFocalNext	6.95	7.76
YOLOv5sNADHFOCAL	6.95	7.76
YOLOFM	3.63	5.78
YOLOFM-SPPF-LLR	11.63	15.29
YOLOFM-Conv	2.86	4.98
YOLOFM-NADH+C3	**3.90**	**6.27**
YOLOFM-FocalNext	5.94	10.14
YOLOFM-Freeze	3.63	5.78

The analysis of the results revealed that YOLOv5s demonstrated outstanding performance, achieving the highest frame rate (87.02 FPS), which makes it particularly well-suited for real-time applications that prioritize speed and detection robustness. With an overall accuracy of 0.92, precision of 0.96, recall of 0.95, and mAP50 of 0.98, YOLOv5s stood out as a reliable and efficient model for fire and smoke detection tasks. On the other hand, YOLOFM-NADH+C3 also showcased competitive performance, matching YOLOv5s in overall accuracy (0.92), while excelling in computational efficiency with only 3.90 million parameters and 6.27 GFLOPs. This lightweight configuration makes YOLOFM-NADH+C3 a strong candidate for resource-constrained environments, particularly in applications where precise localization is critical. These results highlight the strengths of both models in different operational contexts, with YOLOv5s excelling in speed-critical scenarios and YOLOFM-NADH+C3 offering a balance between accuracy and computational efficiency.

Based on these findings, YOLOv5s was selected for scenarios prioritizing speed with high overall accuracy, while YOLOFM-NADH+C3 was chosen for tasks requiring high overall accuracy with lower computation cost. These models were further tested in voting-based fire and smoke detection and 3D localization experiments, as detailed in subsequent sections.

Voting-Based Fire & Smoke Detection and 3D Localization

In this section, we evaluate the system's performance in detecting and localizing fire and smoke using a voting-based decision mechanism combined with 3D localization. The system integrates three strategically placed cameras, each processing video frames through either YOLOv5s or YOLOFM-NADH+C3. For every detected object, the system provides the bounding box coordinates, detection confidence, and projected 3D position (x, y, z). A confidence threshold of 0.5 is applied to determine valid detections. When at least two cameras report detections with confidence values above this threshold, the system votes "Fire Detected" or "Smoke Detected" or both and calculates the average 3D coordinates from the valid detections. If fewer than two cameras meet the threshold, the system outputs "No Fire And Smoke Detected."

A confidence threshold of 0.5 is applied to classify valid detections. If at least two cameras report confidence values exceeding the threshold, the system votes to confirm the detection as "Fire Detected," "Smoke Detected," or both. The system then calculates the average 3D coordinates of the valid detections to localize the object. If fewer than two cameras meet the threshold, the system outputs "No Fire And Smoke Detected."

The voting mechanism ensures robustness by reducing the impact of errors or misdetections from individual cameras. Each camera contributes independently to the voting process, and the final decision is made based on the majority vote. For 3D localization, the x, y, and z coordinates are averaged to achieve a reliable representation of the detected object's position. The z-coordinate (representing vertical height) is determined using trigonometric calculations based on the camera’s position and viewing parameters.

In Figure 5 (a), the results of detection from the YOLOv5s model demonstrate the system's ability to detect and localize real fire accurately. All three cameras reported high-confidence detections, resulting in calculated 3D coordinates that were closely aligned with the ground truth. The predicted location for YOLOv5s was x: 20.4 cm, y: 20.7 cm, z: 12.2 cm, yielding errors of approximately 0.4 cm in x, 0.7 cm in y, and 1.3 cm in z. Using Eq. (13), the overall localization error was determined to be approximately 1.53 cm, highlighting the model's precision. Figure 5 (b) illustrates the system's response to a fake fire scenario, where the YOLOv5s model detected the fake fire, but the system successfully withheld a "Fire Detected" vote due to insufficient confidence levels (below 0.5) from the cameras. This outcome emphasizes the system's reliability in avoiding false positives.

(a) (b)

Figure 5. (a) YOLOv5s Fire Detected (b) YOLOv5s On Fake Fire

Similarly, in Figure 6 (a), the YOLOFM-NADH+C3 model demonstrates its capability to detect and localize real fire with high accuracy. The predicted location for the fire was x: 20.0 cm, y: 20.2 cm, z: 13.4 cm, resulting in errors of 0.0 cm in x, 0.2 cm in y, and 0.1 cm in z. Using Eq. (13), the overall localization error was calculated to be approximately 0.22 cm, showcasing the model's exceptional spatial accuracy. Figure 6 (b) presents the model's performance on a fake fire scenario, where the system correctly classified the situation as "No Fire and Smoke Detected" due to insufficient confidence votes from the cameras, despite the model detecting the fake fire.

(a) (b)

Figure 6. (a) YOLOFM-NADH+C3 Fire Detected (b) YOLOFM-NADH+C3 On Fake Fire

These experiments demonstrate the robustness and reliability of the system in distinguishing real fire from fake fire under varying conditions. While this evaluation focused on fire detection, the system is also capable of detecting smoke using the same voting mechanism and 3D localization framework. This flexibility underscores the system's potential as a comprehensive solution for fire and smoke detection in real-time applications, as evidenced by its consistent and accurate results. evidenced by its consistent and accurate results.

D. Conclusion

In this study, we developed and evaluated a fire detection and localization system that integrates object detection models, a confidence-based voting mechanism, and 3D localization. By leveraging three strategically placed cameras, the system calculates the 3D coordinates (x,y,z) of detected fire or smoke, ensuring precise spatial awareness. A robust voting mechanism requires at least two cameras to validate a detection with a confidence score above 0.5, significantly reducing false positives and improving detection reliability. The 3D localization process demonstrated high accuracy across experiments, effectively identifying fire locations with minimal error. In real fire scenarios, the system localized fire with YOLOv5s achieving an average localization error of 1.53 cm and YOLOFM-NADH+C3 achieving an even smaller error of 0.22 cm, underscoring its superior spatial precision. In fake fire scenarios, the system consistently withheld a "fire detected" classification when confidence votes were insufficient, showcasing its robustness against false alarms. The system testing phase compared the performance of two models: YOLOv5s and YOLOFM-NADH+C3. YOLOv5s achieved an overall accuracy of 92%, with a precision of 96%, recall of 95%, and mAP50 of 98%, alongside an impressive frame rate of 87.02 FPS, making it well-suited for real-time applications. On the other hand, YOLOFM-NADH+C3 matched YOLOv5s in overall accuracy and detection metrics while delivering superior localization precision, albeit with a slightly lower FPS of 54.82. This highlights a trade-off between speed and spatial accuracy, with YOLOFM-NADH+C3 excelling in precision-critical scenarios and YOLOv5s performing better in high-speed environments. Overall, the system proved effective in detecting and localizing fire, with the integration of voting and 3D localization ensuring reliable results even in challenging scenarios. Future work could focus on improving camera calibration for even greater spatial accuracy. This study highlights the potential of voting-based 3D localization systems to provide accurate, reliable fire and smoke detection for real-time safety applications.

E. Acknowledgment

The authors would like to express their heartfelt gratitude to Universitas Ciputra Surabaya for providing research funding under the Program Dana Internal Penelitian Skema DIP Mahasiswa Tahun Anggaran 2024/2025, in accordance with the agreement between LPPM Universitas Ciputra Surabaya and the researchers, No. 017/UC-LPPM/DIP/SP3H/IX/2024, dated 20 September 2024. The authors also extend their sincere thanks to the authors of the YOLOFM Journal for generously providing the datasets and the model utilized in this research.

F. References

[1] F. Khan, Z. Xu, J. Sun, F. M. Khan, A. Ahmed, and Y. Zhao, “Recent Advances in Sensors for Fire Detection,” *Sensors*, vol. 22, no. 9, p. 3310, Apr. 2022, doi: https://doi.org/10.3390/s22093310. [Accessed: 08-November-2024]

[2] L. Terrei *et al.*, “Experimental and numerical multi-scale study of spruce wood degradation under inert atmosphere,” *Fire Safety Journal*, vol. 130, p. 103598, Jun. 2022, doi: https://doi.org/10.1016/j.firesaf.2022.103598. [Accessed: 08-November-2024]

[3] H. Office, “Fire and rescue incident statistics: England, year ending June 2024,” *GOV.UK*, Oct. 24, 2024. Available: https://www.gov.uk/government/statistics/fire-and-rescue-incident-

statistics-england-year-ending-june-2024/fire-and-rescue-incident-statistics-england-year-ending-june-2024. [Accessed: 08-November-2024]

[4] J. Pincott, Paige Wenbin Tien, S. Wei, and John Kaiser Calautit, "Development and evaluation of a vision-based transfer learning approach for indoor fire and smoke detection," *Building Services Engineering Research and Technology*, vol. 43, no. 3, pp. 319–332, Mar. 2022, doi: https://doi.org/10.1177/01436244221089445. [Accessed: 08-November-2024]

[5] M. Hassaballah and A. I. Awad, "Deep Learning in Computer Vision," Mar. 2020, doi: https://doi.org/10.1201/9781351003827.

[6] A. A. Mohammed, Nagur Babu Alapaka, Chaitanya Gudivada, K. P. Bharath, and M. R. Kumar, "Computer Vision Based Autonomous Fire Detection and IoT Based Fire Response System," *Algorithms for intelligent systems*, pp. 551–560, Jan. 2021, doi: https://doi.org/10.1007/978-981-16-3246-4_43. [Accessed: 08-November-2024]

[7] M. Chetoui and M. A. Akhloufi, "Fire and Smoke Detection Using Fine-Tuned YOLOv8 and YOLOv7 Deep Models," *Fire*, vol. 7, no. 4, pp. 135–135, Apr. 2024, doi: https://doi.org/10.3390/fire7040135. [Accessed: 08-November-2024]

[8] M. Yaseen, "What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector," *arXiv.org*, 2024. https://arxiv.org/abs/2408.15857. [Accessed: 08-November-2024]

[9] S. Saponara, A. Elhanashi, and A. Gagliardi, "Real-time video fire/smoke detection based on CNN in antifire surveillance systems," *Journal of Real-Time Image Processing*, Nov. 2020, doi: https://doi.org/10.1007/s11554-020-01044-0. [Accessed: 08-November-2024]

[10] U. Nepal and H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs," *Sensors*, vol. 22, no. 2, p. 464, Jan. 2022, doi: https://doi.org/10.3390/s22020464. [Accessed: 08-November-2024]

[11] Tautvydas Kvietkauskas, E. Pavlov, Pavel Stefanovič, and Birutė Pliuskuvienė, "The Efficiency of YOLOv5 Models in the Detection of Similar Construction Details," *Applied Sciences*, vol. 14, no. 9, pp. 3946–3946, May 2024, doi: https://doi.org/10.3390/app14093946. [Accessed: 09-November-2024]

[12] X. Geng, Y. Su, X. Cao, H. Li, and L. Liu, "YOLOFM: an improved fire and smoke object detection algorithm based on YOLOv5n," *Scientific Reports*, vol. 14, no. 1, Feb. 2024, doi: https://doi.org/10.1038/s41598-024-55232-0. [Accessed: 12-June-2024]

[13] Temitope Ibrahim Amosa *et al.*, "Multi-Camera Multi-Object Tracking: A Review of Current Trends and Future Advances," *Neurocomputing*, vol. 552, pp. 126558–126558, Oct. 2023, doi: https://doi.org/10.1016/j.neucom.2023.126558. [Accessed: 10-November-2024]

[14] P. Kadam, G. Fang, and Ju Jia Zou, "Object Tracking Using Computer Vision: A Review," *Computers*, vol. 13, no. 6, pp. 136–136, May 2024, doi: https://doi.org/10.3390/computers13060136. [Accessed: 10-November-2024]

[15] S. Matsuda, N. Techasarntikul, and H. Shimonishi, “Multi-Camera 3D Position Estimation using Conditional Random Field,” *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 1900–1908, Oct. 2023, doi: https://doi.org/10.1109/iccvw60793.2023.00205. [Accessed: 13-June-2024]

[16] H. Yamaguchi, K. Tateno, and K. Horio, "3D position estimation of multiple fish in a small tank using DeepLabCut and shortest distance calculation between two straight lines," *ICIC Express Letters*, vol. 18, no. 8, pp. 835–842, 2024. [Online]. Available: https://doi.org/10.24507/icicel.18.08.835. [Accessed: 11-November-2024]

[17] H. Xu, B. Li, and F. Zhong, “Light-YOLOv5: A Lightweight Algorithm for Improved YOLOv5 in Complex Fire Scenarios,” Aug. 2022, doi: https://doi.org/10.48550/arxiv.2208.13422. [Accessed: 11-November-2024]