Privacy-aware smart camera for abnormal event detection in home environments

Doan Huong Giang[1], Ho Anh Dung[2*]

[1]Faculty of Control and Automation, Electric Power University, 235 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam;
[2]Department of Information Technology, East Asia University of Technology, Phan Tay Nhac, Xuan Phuong, Hanoi, Vietnam.
[*]Corresponding author: dungha@eaut.edu.vn
Received 28 Nov. 2025; Revised 26 Jan. 2026; Accepted 10 Feb. 2026; Published 25 Feb. 2026.
DOI: https://doi.org/10.54939/1859-1043.j.mst.109.2026.137-145

ABSTRACT

Ensuring both effective monitoring and user privacy is essential in home surveillance applications. This study proposes a privacy-aware smart camera system integrating a servo-controlled mechanical housing with a dual-branch deep learning framework. That is a smart camera housing equipped with servo-controlled lids and dual operation modes, local button control and WiFi-based remote control, providing convenient usage while preventing unintended image capture. To detect hazardous household events, we constructed the EPUabInhouse dataset and proposed a dual-branch framework that integrates YOLOv8 for spatial analysis with RAFT optical flow for motion representation. Experimental results show that incorporating RAFT leads to a relative improvement of 2.02% to 4.15% in F1-score across different classes and significantly reduces background misclassification. These enhancements demonstrate the effectiveness and practical applicability of the proposed privacy-aware home surveillance system.

Keywords: Convolutional neural network; Deep learning; Optical flow; Anomaly detection; Smart camera.

1. INTRODUCTION

Home safety monitoring has become an increasingly critical need, as incidents such as fire, smoke, electrical short circuits, sudden falls, or abnormal behaviors can occur at any time and cause serious damage. Recent studies have confirmed that computer-vision-based surveillance systems can support early detection, mitigate risks, and enhance safety levels in living environments [1-4]. However, conventional camera systems typically record continuously, raising concerns about privacy as well as the risk of personal data leakage in household settings [5].

In addition to privacy and security issues, abnormal event detection in indoor environments remains challenging due to complex backgrounds, varying illumination conditions, diverse object shapes, and complicated motion patterns [3, 6]. Studies on fire/smoke detection [1-4], fall detection [7, 8], and video-based anomaly recognition [6] have all shown that detection accuracy strongly depends on the model's ability to jointly exploit spatial and temporal features. Nevertheless, most existing systems focus primarily on algorithmic (software) solutions and lack hardware designs tailored to home environments, where ease of use and privacy protection are essential.

To address these challenges, this study proposes a smart camera system integrated into a servo-driven mechanical housing with shutter lids that can be opened/closed either via physical push buttons or WiFi-based remote control, allowing users to actively control the recording process. This design ensures that the camera only operates when truly needed, thereby enhancing privacy in daily activities. In addition, the system is equipped with an uninterruptible power supply (UPS) to maintain stable operation even in the event of sudden power outages. In parallel with the hardware design, we construct the EPUabInhouse dataset, which includes common hazardous situations in home environments. Based on this dataset, we develop a dual-branch model that combines YOLO[9] and RAFT[10] to simultaneously exploit spatial features and optical-flow

dynamics. This combination enables the model to better recognize events with diffuse or soft-boundary characteristics, such as smoke, as well as to more clearly distinguish dangerous behaviors such as falls or unsafe interactions with electrical outlets.

The main contributions of this paper are as follows: (1) the design of a smart camera housing that provides flexible, user-friendly control while protecting user privacy; (2) the construction of the EPUabInhouse dataset containing diverse types of household incidents; (3) the proposal of a YOLO–RAFT dual-branch model that fuses spatial and temporal information to improve the reliability of abnormal event detection; and (4) experimental validation demonstrating that the combined model achieves superior performance compared to standalone YOLO and significantly reduces background misclassification.

The remainder of this paper is organized as follows. Section 2 describes the proposed solution in detail. Experimental results are presented in section 3. Section 4 concludes the paper and discusses directions for future work.

2. PROPOSED METHOD

2.1. Design of the camera housing

Figure 1 illustrates the mechanical design of the smart camera housing installed at the intersection of two walls. The technical drawings detail its 100 mm × 100 mm × 120 mm structure, providing space for the camera and internal actuators. Figures 1(a) and 1(b) show the enclosure layout, where position (1) mounts the camera centrally for an optimal field of view, and positions (2), (3), and (4) house the servo motors that drive the protective lids.

Figure 1(c) presents the completed prototype. The exterior appears compact and streamlined, while the interior reveals the centrally mounted camera and three symmetrically arranged servos controlling the shutter panels. This design allows the camera to be fully concealed when needed, thereby enhancing user privacy and preventing unintended image capture or data leakage during occupied periods.

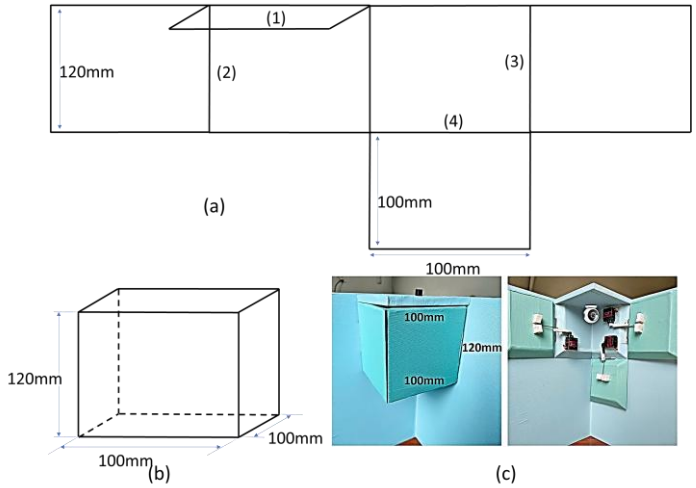


Figure 1. *Hardware design of the smart camera enclosure.*

The electronic control system of the camera housing is developed based on the overall block diagram shown in figure 2, with the detailed circuit schematic presented in figure 3. The integration of mechanical and electronic design ensures stable, reliable, and autonomous system operation. Figure 2 presents the block diagram of the proposed IoT-based smart camera system. The ESP32 serves as the central controller, interfacing with the camera module, servo motors, user buttons, LCD display, and alert components. Local control is achieved through physical buttons, while

remote control is enabled via WiFi communication. The power and UPS module ensures uninterrupted operation in case of a power failure.

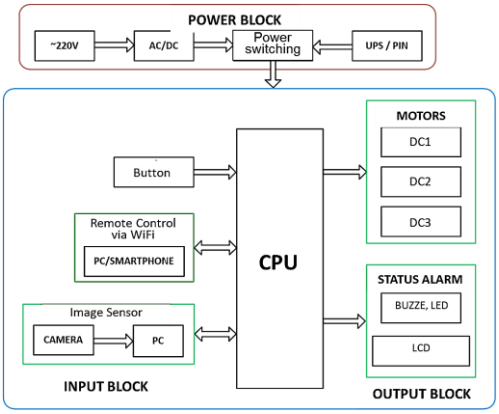


Figure 2. *Block diagram of the smart camera housing control system.*

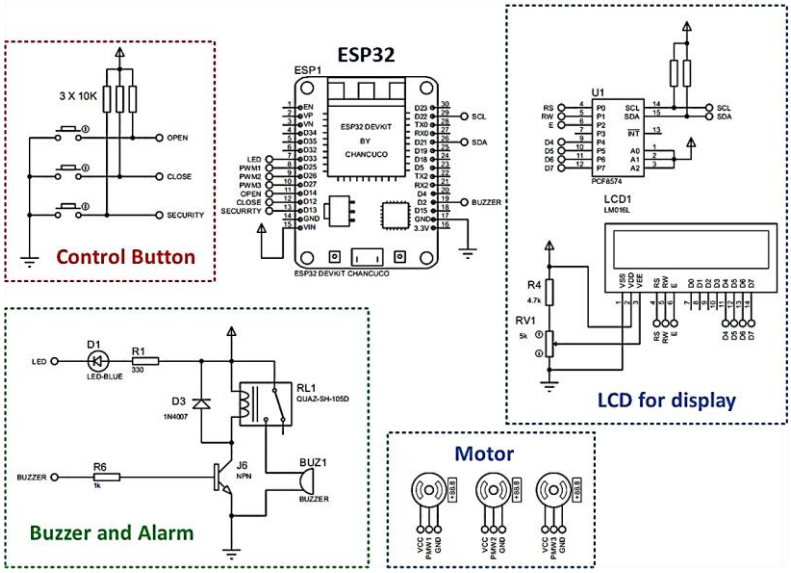


Figure 3. *Detailed circuit schematic of the smart camera housing controller.*

Figure 3 shows the controller circuit of the smart camera housing. The ESP32 processes inputs from physical buttons and WiFi commands to control the servo-driven privacy shutters, update system status via an LCD module, and trigger buzzer and LED alerts, ensuring reliable operation under both local and remote control. In addition, the system supports remote control via a mobile application using the ESP32's built-in WiFi module. Through the app, users can send commands, monitor the housing status in real time, and activate security functions without physical access. The combination of both control modes offers greater convenience, safety, and privacy protection, making the system well-suited for home surveillance applications.

All signals from both control mechanisms are processed by the ESP32, which drives the servo motors, updates the LCD via the PCF8574 module, and triggers the buzzer and LED when abnormal access is detected. This integration of electromechanical control, network communication, and user interaction ensures smooth operation, safety, and effective privacy protection.

Figure 4 illustrates the design of the automatic power-switching (UPS) system used to maintain continuous operation of the camera housing during power outages. In this design, the 220 V AC

source is stepped down by transformer T1 and rectified by bridge B1 to generate a 12 V DC supply. The rectified voltage is filtered by capacitor C1 and protected by fuse F1 before being distributed to two branches: one providing 12 VDC directly to the load, and the other regulated by the 7805 (U1) to produce a stable 5 VDC output. To ensure uninterrupted power, a 12 V backup battery is connected through diodes D2 and D3 in an ORing configuration, enabling automatic switchover when mains power is lost. Diode D1 prevents reverse current flow, while capacitors C2 and C3 filter noise at the regulator's input and output. LED D4 and resistor R2 provide a visual indication of the power status. This configuration ensures stable 12 VDC and 5 VDC supplies for the system's control, sensing, and communication modules.

Figure 4 illustrates the schematic of the automatic power-switching (UPS) system designed to ensure the continuous operation of the smart camera housing. Under normal conditions, the AC mains supply is rectified and regulated to provide stable 12 VDC and 5 VDC outputs for both medium-power loads and low-voltage control modules. During a power outage, a backup battery is automatically connected through a diode-based ORing mechanism, enabling seamless power switchover without interrupting system operation. Compared with the design in [1], which adopts a simpler linear-regulated structure with basic diode switching and a single 12 V output, the proposed UPS incorporates dual fuses, multi-stage filtering, and enhanced reverse-current isolation. These improvements increase operational stability and safety under fluctuating power conditions while supporting multi-level power delivery, making the system more suitable for smart electromechanical applications.

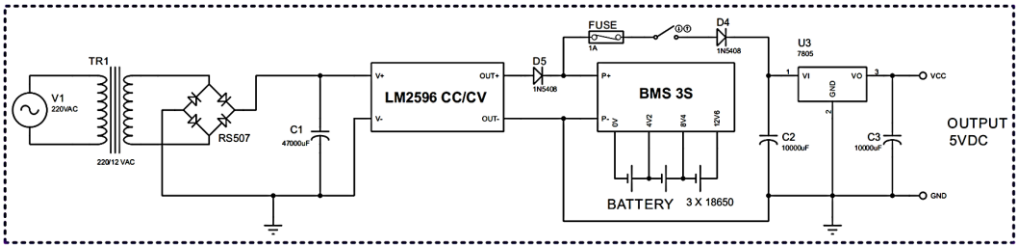


Figure 4. *Design schematic of the automatic power-switching (UPS) system.*

2.2. Abnormal event detection model for home environments

2.2.1. Abnormal event dataset for home environments

In this study, we construct an abnormal event dataset for home environments with three labels: electrical misuse, fire/smoke, and fall. The dataset combines our self-recorded videos with additional images and videos collected from the Internet, resulting in a diverse set of household incident scenarios. We name this dataset EPUabInhouse. Figure 5 shows sample images of smoke and fire from the dataset.

Figure 5. *Illustration images of the EPUabInhouse dataset.*

This dataset contains images of the most common hazardous events that may occur in household indoor environments, captured under complex backgrounds, varying lighting conditions, and at different times and locations. The EPUabInhouse dataset includes both videos and images. For self-recorded data, we use a HiK-Vision DS-2CD2643G2-IZS camera, which

provides RGB frames at 30 fps with a resolution of 2688 × 1520 pixels. Each recorded video ranges from 1 to 10 minutes. The Internet-sourced portion consists of videos and images with various resolutions. After collection, all data are converted into image frames, sampled at 3 fps, and manually annotated using LabelImg with the four labels listed in table 1.

Table 1. *Description of the EPUabInhouse dataset.*

No.	Incident type	Number of Images
1	Electrical Misuse (EM)	5293
2	Fall (F)	5203
3	Fire (FS)	5478
4	No Incident (NI)	11143

For the EPUabInhouse dataset, we annotate image regions corresponding to each incident type using the label indices defined in table 1, where the Fire label comprises fire or smoke. This dataset is subsequently used to evaluate the proposed image-based abnormal event detection model for home environments. The EPUabInhouse dataset is currently private and will be made available upon reasonable request for research purposes.

2.2.2. Image-based abnormal event detection

In the image-based abnormal event detection module, we employ a model that integrates YOLO with optical flow, as illustrated in figure 6.

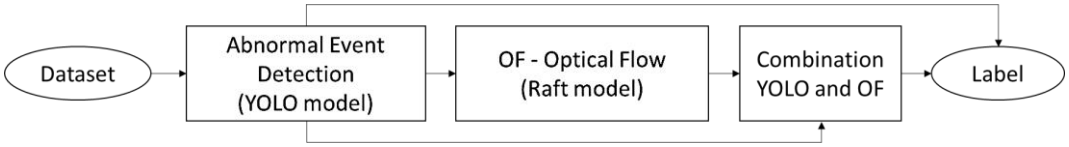


Figure 6. *Diagram of the abnormal event detection system.*

Figure 6 illustrates the dual-branch architecture used to analyze abnormal events in surveillance video sequences. The first branch performs abnormal event detection using the YOLO model, operating in the spatial domain. The second branch computes optical flow using the RAFT model, capturing motion information in the temporal domain. Combining these two modalities improves the system's reliability in detecting undesired phenomena such as smoke, fire, or abnormal human activities. In the event-detection branch, YOLO processes a single frame $I_t \in R^{W\times H\times 3}$ and produces a set of predictions as formulated in equation (1):

$$B_t = \{(b_i, c_i, s_i) | i = 1, \dots, N_t\} \quad (1)$$

In this formulation, $b_i = (x_i, y_i, w_i, h_i)$ denotes the predicted bounding box, c_i is the class label, and $s_i \in [0,1]$ represents the confidence score. The model is trained using a multi-component loss function that includes localization loss, classification loss, and confidence loss, as expressed in equation (2). This objective enables the model to learn both the geometric structure and the appearance characteristics of abnormal entities within the scene.

$$L_{YOLO} = \gamma_{coor} L_{bbox} + \gamma_{obj} L_{obj} + \gamma_{cls} L_{cls} \quad (2)$$

In parallel, the RAFT model estimates the motion field between two consecutive frames I_t _tIt and I_{t+1}. The resulting optical flow field is defined as in equation (3):

$$F_{t\to t+1} = M_{RAFT}(I_t, I_{t+1}) = (u(x,y), v(x,y)) \quad (3)$$

Where, u and v represent the horizontal and vertical displacements at each pixel (x,y), respectively. RAFT extracts features from both frames and constructs a global correlation volume to produce accurate and stable optical flow estimates, an important capability for tracking smoke propagation or soft-boundary abnormal motions. The correlation between feature vectors $\emptyset(I_t)_i$ and $\emptyset(I_{t+1})_j$ is computed using their dot product, as expressed in equation (4):

$$C(i,j) = \langle \emptyset(I_t)_i, \emptyset(I_{t+1})_j \rangle \quad (4)$$

Here, $C(i,j)$ denotes the global correlation volume between two consecutive frames. Each element $C(i,j)$ measures the similarity between the i^{th} feature in I_t and the j^{th} feature in I_{t+1}. This 4D volume spans all possible feature correspondences between the two images, which is crucial for RAFT to accurately identify matching points—especially under large or complex motions. The operator $\emptyset(.)$ represents the feature extractor, where $\emptyset(I) = f \in R^d$ and d is the dimensionality of the RAFT feature vector. The detailed computation of $C(i,j)$ is given in equation (5):

$$C(i,j) = \langle f_i, f_j \rangle = \sum_{k=0}^{d-1} f_i^k f_j^k \tag{5}$$

The integration of the two processing branches offers significant advantages. The YOLO model provides spatial information about objects and abnormal regions, while RAFT contributes motion dynamics through optical flow variation. As a result, phenomena such as smoke, characterized by diffuse and ambiguous boundaries, can be captured through abnormal increases in the optical flow gradient. In contrast, for fall events where a person becomes immobilized (e.g., due to injury or fainting), the optical flow typically exhibits minimal change after the fall. In this study, we quantify these motion patterns using the optical flow gradient, defined in equation (6):

$$M_t = \|\nabla F_t\|_2 = \sqrt{\left(\frac{\partial u_t}{\partial x}\right)^2 + \left(\frac{\partial u_t}{\partial y}\right)^2 + \left(\frac{\partial v_t}{\partial x}\right)^2 + \left(\frac{\partial v_t}{\partial y}\right)^2} \tag{6}$$

Combining this motion cue with the spatial information from YOLO enables more reliable decision-making. Consequently, the dual-branch architecture enhances the system's sensitivity to complex abnormal events, particularly in environments with significant illumination changes or when abnormal objects exhibit unclear or diffuse boundaries.

2.2.3. Metric and protocol evaluation

Metric: The model's performance is evaluated using Precision, Recall, F1-score, and mAP (Mean Average Precision). Precision measures the proportion of correct positive predictions, while Recall reflects the proportion of correctly detected incident regions relative to all actual incident regions. F1-score represents the harmonic mean of Precision and Recall. The Average Precision (AP) is computed for each class, and mAP is obtained by averaging AP across all classes. Detailed evaluation results using these metrics are presented in section 3.

Protocol evaluation: In this paper, we employ a k-fold cross-validation protocol [1] on the EPUabInhouse dataset, which is divided into ten equal parts. In each iteration, eight parts are used for training, one part for validation, and the remaining part for testing. Each iteration produces corresponding Precision, Recall, and mAP values. The overall performance of the model is obtained by averaging these metrics across all ten folds. This evaluation procedure provides a reliable assessment of the model's stability and generalization capability over the entire dataset.

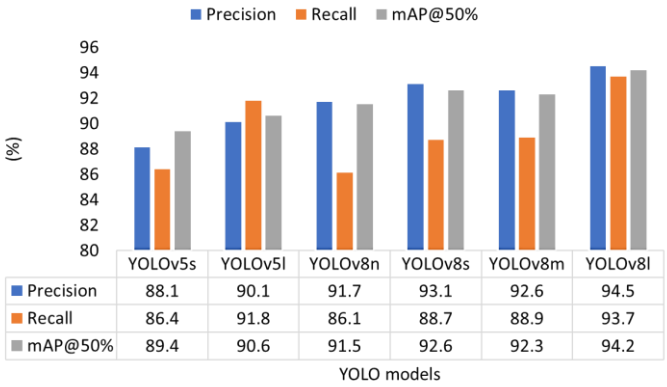
3. EXPERIMENTAL RESULTS

The experiments are conducted on a computer equipped with an Intel Core i5-11400H CPU, an NVIDIA GeForce GTX 1650 GPU, and 8 GB of RAM, using Python as the programming environment. The model is trained with a batch size of 64, a learning rate of 0.0001, and 100 epochs. Evaluations are performed on multiple YOLO variants, including YOLOv5s[11], YOLOv6[12], YOLOv7[13], YOLOv8n, YOLOv8s, and YOLOv8m[14]. The experiments assess: (1) the abnormal event detection performance of these YOLO versions, and (2) the effectiveness of the dual-branch architecture when integrating YOLO and RAFT[10].

3.1. Abnormal event detection performance of YOLO models

In this section, we independently evaluate the performance of each YOLO variant on the

EPUabInhouse abnormal event dataset. The models are assessed using three standard object-detection metrics: Precision, Recall, and mAP. The results are presented in figure 7.



	YOLOv5s	YOLOv5l	YOLOv8n	YOLOv8s	YOLOv8m	YOLOv8l
Precision	88.1	90.1	91.7	93.1	92.6	94.5
Recall	86.4	91.8	86.1	88.7	88.9	93.7
mAP@50%	89.4	90.6	91.5	92.6	92.3	94.2

Figure 7. *Abnormal event detection performance of YOLO models.*

Figure 7 summarizes the Precision, Recall, and mAP@50% scores of six YOLO models: YOLOv5s, YOLOv5l, YOLOv8n, YOLOv8s, YOLOv8m, and YOLOv8l, evaluated on four incident classes: electrical misuse, fall, fire/smoke, and no incident. The results indicate that:

- In terms of Precision, YOLOv8l achieves the highest score (94.5%), followed by YOLOv8s (93.1%) and YOLOv8m (92.6%). Overall, the YOLOv8 family demonstrates stronger classification capability compared to YOLOv5, whose models (YOLOv5s and YOLOv5l) obtain lower Precision scores (88.1% and 90.1%), reflecting architectural differences in feature extraction.

- For Recall, YOLOv8l again leads with 93.7%, indicating stable detection performance on incident-containing frames. YOLOv5l also performs well (91.8%), while YOLOv5s, YOLOv8n, YOLOv8s, and YOLOv8m range between 86.1% and 88.9%.

- Regarding mAP@50, YOLOv8l attains the best result (94.2%), followed by YOLOv8s (92.6%) and YOLOv8m (92.3%). As expected, YOLOv5s and YOLOv5l achieve lower mAP@50 scores (89.4% and 90.6%) due to their earlier-generation architecture.

Overall, these results highlight the superior performance of YOLOv8l among the evaluated models, establishing it as a strong baseline for subsequent integration with RAFT in the dual-branch detection framework.

3.2. Detection accuracy of the combined YOLO–RAFT framework

In this experiment, each incident category is evaluated using 10 short indoor video clips, with durations ranging from 10 seconds to 1 minute, collected by our team. Each video is sampled at 3 frames per second, and every sampled frame is annotated according to the incident types defined in table 1. The evaluation is then performed on the entire set of sampled frames. Table 2 presents the results obtained from the YOLOv8l model and the combined YOLOv8l–RAFT model, reporting Precision, Recall, and F1-score for both approaches.

The results show that YOLOv8l+RAFT consistently outperforms the standalone YOLOv8l across all classes. In the original YOLOv8l model, the “Fall” and “Fire” classes achieve F1-scores of around 92%, while the “No Incident” class is lower due to background confusion. After integrating RAFT, all classes improve markedly, with F1-scores of 95 - 96%. The “No Incident” class increases from 91.11% to 95.26%, demonstrating RAFT’s ability to reduce motion noise and suppress false alarms. Figure 8 illustrates the confusion matrices of both models. All confusion matrices are averaged over 10-fold cross-validation.

Table 2. *Experimental results of YOLOv8l combined with optical flow (RAFT).*

Incident type	YOLOv8l			YOLOv8l+RAFT		
	Precision (%)	**Recall (%)**	**F1-score (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Electrical misuse	94.92	93.5	94.21	96.67	95.8	96.23
Fall	90.97	93.7	92.32	95.14	96.0	95.57
Fire	91.54	93.0	92.26	95.59	95.3	95.44
No incident	92.57	89.7	91.11	95.11	95.4	95.26

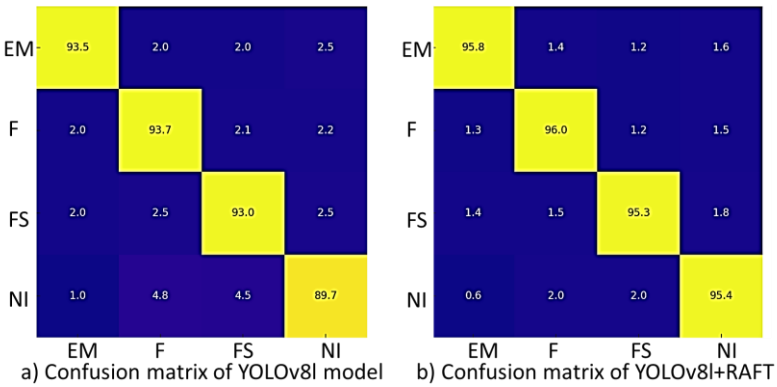


Figure 8. *Confusion matrix of the EPUabInhouse dataset.*

The confusion matrix analysis further confirms these improvements. Correct recognition of the three incident classes rises to 95.3 - 96.0%, compared to 93.0 - 93.7% with YOLOv8l. The largest gain is again seen in the "No Incident" class, where accuracy increases from 89.7% to 95.4%, and misclassification rates drop from 4 - 5% to about 2%. Cross-class confusion also decreases (from 2.0 - 2.5% to 1.0 - 1.5%), indicating better separation of motion patterns. Thus, integrating RAFT significantly enhances robustness, reduces background misclassification, and improves the reliability of abnormal event detection in home environments.

Overall, the confusion matrix results confirm that integrating RAFT significantly improves detection accuracy, reduces background misclassification, and enhances the model's stability in recognizing both abnormal events and non-event frames. These findings provide clear evidence of the effectiveness of combining spatial and temporal information for video-based anomaly detection.

4. CONCLUSIONS

The study developed a smart camera housing equipped with servo-controlled lids, supporting both manual and WiFi-based operation to ensure convenient use and protect user privacy by fully covering the camera when needed. Using the EPUabInhouse dataset, we proposed a dual-branch YOLO–RAFT framework that integrates spatial and motion information. Experimental results show that YOLOv8l+RAFT significantly outperforms YOLOv8l, raising the F1-score to 95% and 96% and substantially reducing background misclassification, thereby demonstrating the effectiveness of the proposed approach.

REFERENCES

[1]. Hồ Anh Dũng, Đoàn Thị Hương Giang, Trần Đình Hùng, Ma Khánh Tùng, Nguyễn Huyền Tiến An, Bùi Thị Duyên, *"Hệ thống phát hiện khói và cháy thông minh đa thể thức"*, Tạp chí Khoa học và Kỹ thuật – Học viện Kỹ thuật Quân sự, vol. 97, no. 97, pp. 138–147, (2024) (in Vietnamese).

[2]. Elhanashi, S. Essahraui, P. Dini, S. Saponara, *"Early Fire and Smoke Detection Using Deep Learning: A Comprehensive Review"*, Applied Sciences, vol. 15, no. 18, article 10255, pp. 1–34, (2025).

doi:10.3390/app151810255.
[3]. D. Gragnaniello, A. Greco, C. Sansone, B. Vento, *"Fire and Smoke Detection from Videos: A Literature Review under a Novel Taxonomy"*, Expert Systems with Applications, vol. 260, article 124783, pp. 1–32, (2024). doi:10.1016/j.eswa.2024.124783.
[4]. V. Carletti, A. Greco, A. Saggese, B. Vento, *"A Smart Visual Sensor for Smoke Detection Based on Deep Neural Networks"*, Sensors, vol. 24, article 4519, (2024). doi:10.3390/s24144519.
[5]. K.-S. Wong, N. A. Tu, A. Maratkhan, M. F. Demirci, *"A Privacy-Preserving Framework for Surveillance Systems"*, Proceedings of the 10th International Conference on Communication and Network Security (ICCNS), Tokyo, Japan, pp. 91–98, (2021). doi:10.1145/3442520.3442524.
[6]. R. K. Yadav, R. Kumar, *"A Survey on Video Anomaly Detection"*, IEEE Delhi Section Conference (DELCON), New Delhi, India, pp. 1–5, (2022). doi:10.1109/DELCON54057.2022.9753580.
[7]. E. Alam, A. Sufian, P. Dutta, M. Leo, *"Vision-based Human Fall Detection Systems Using Deep Learning: A Review"*, Computers in Biology and Medicine, vol. 146, article 105626, (2022). doi:10.1016/j.compbiomed.2022.105626.
[8]. C. Li, M. Liu, X. Yan, G. Teng, *"Research on CNN-BiLSTM Fall Detection Algorithm Based on Improved Attention Mechanism"*, Applied Sciences, vol. 12, article 9671, (2022). doi:10.3390/app12199671.
[9]. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, *"You Only Look Once: Unified, Real-Time Object Detection"*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788, (2016). doi:10.1109/CVPR.2016.91.
[10]. Z. Teed, J. Deng, *"RAFT: Recurrent All-Pairs Field Transforms for Optical Flow"*, European Conference on Computer Vision (ECCV), LNCS 12347, pp. 402–419, Springer, (2020). doi:10.1007/978-3-030-58536-5_24.
[11]. G. Jocher et al., *"YOLOv5 by Ultralytics"*, GitHub Repository, (2020). Available: https://github.com/ultralytics/yolov5
[12]. C. Li et al., *"YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications"*, arXiv preprint arXiv:2209.02976, (2022).
[13]. C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, *"YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors"*, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, pp. 7464–7475, (2023).
[14]. G. Jocher et al., *"Ultralytics YOLOv8 Technical Report"*, Ultralytics, pp. 1–45, (2023).

TÓM TẮT

Camera thông minh bảo vệ quyền riêng tư cho phát hiện sự kiện bất thường trong hộ gia đình

Giám sát an toàn trong hộ gia đình đòi hỏi sự cân bằng giữa hiệu quả phát hiện và bảo vệ quyền riêng tư. Nghiên cứu này trình bày một hộp camera thông minh với cơ cấu nắp che điều khiển bằng servo, hỗ trợ cả nút bấm trực tiếp và điều khiển qua WiFi, giúp việc sử dụng thuận tiện và đảm bảo không ghi hình ngoài ý muốn. Để phát hiện các sự kiện nguy hiểm, chúng tôi xây dựng bộ dữ liệu EPUabInhouse và đề xuất mô hình hai nhánh kết hợp YOLOv8 với luồng quang học RAFT nhằm tận dụng đồng thời thông tin không gian và chuyển động. Kết quả thực nghiệm cho thấy mô hình tích hợp giúp tăng F1-score thực tế từ 2.02% đến 4.15% tùy lớp, đồng thời giảm đáng kể các trường hợp nhầm lẫn nền. Các cải thiện này chứng minh tính hiệu quả và khả năng ứng dụng thực tế của hệ thống giám sát gia đình đề xuất.

Từ khoá: Mạng nơ ron tích chập; Học sâu; Luồng quang học; Phát hiện bất thường; Camera thông minh.