Lightweight Anomaly Detection Mechanism based on Machine Learning Using Low-Cost Surveillance Cameras

Yeon-Ji Lee[1[0000-0002-0482-2381]], Na-Eun Park[1[0009-0008-0939-5762]] and Il-Gu Lee[1[0000-0002-5777-4029]]

[1] Sungshin Women's University, Seongbuk-Gu, Seoul, 02844, Korea
{cselab.lyj, nepark.cse, iglee19}@gmail.com

Abstract. As the need for on-site monitoring using surveillance cameras increases, there has been a growing interest in automation research incorporating machine learning. However, traditional research has not resolved the performance and resource efficiency trade-offs. Therefore, we proposed a lightweight learning model that is more efficient and with minimal performance degradation. The proposed model reduces the resolution of the image until the performance is maintained, finding where the trade-off is resolved for each dataset. Using this, we suggested a real-time lightweight fire detection algorithm. The proposed mechanism is approximately 30 times more memory efficient while maintaining the detection performance of traditional methods.

Keywords: surveillance camera, abnormal detection, CNN

1 Introduction

The need for on-site monitoring using surveillance cameras for public management, security, and safety has recently increased. However, interpreting surveillance camera footage is a human task, and as individuals monitor multiple cameras simultaneously, there are clear limitations regarding efficiency and accuracy[1]. When humans manage surveillance cameras, issues arise related to human resources, maintenance costs for installing and managing cameras, and other associated costs[2].

Research on anomaly detection using machine learning is being actively pursued to address these issues. [3-6]. Deep learning, a subset of machine learning, allows training without human intervention and delivers high-level results in object detection, data classification, and natural language processing[3]. In particular, the CNN (convolutional neural networks) model, which directly learns features from datasets, is being utilized for anomaly detection in various fields ranging from medicine to agriculture[4]. However, traditional research has been increasing the resolution of images to the maximum, using ultra-high-resolution images as datasets or relying on high-quality images to enhance the accuracy of CNN models[5]. IoT devices, including surveillance cameras, are constrained in energy, memory, and cost[13-14]. High-quality datasets can maintain high model performance but are unsuitable for real-time surveillance camera detection [6-7]. Various research has been conducted for lightweight CNN learning[8-

11]. However, traditional studies have not effectively addressed the trade-off between model performance and cost, underscoring the need for further research in this area.

The proposed model identifies the optimal resolution where the CNN model can maintain its performance on the fire dataset and suggests a more efficient fire detection mechanism. The suggested mechanism maintained a lowered camera resolution and switched to a clearer quality when the likelihood of fire detection exceeded a threshold. The contributions of this study are as follows.

- By classifying the fire dataset based on the fire size and adjusting the resolution to identify the point at which accuracy is maintained, we have addressed the trade-off issue between performance and cost, a limitation of previous research.
- We proposed a universal mechanism not limited to surveillance cameras, making it easier for lightweight CNN learning to be applied across various research and environments.
- We proposed a lightweight fire detection mechanism that maintains the performance of fire detection while reducing memory consumption by 31.8 times.

The structure of this study is as follows: In Section 2, we investigate and analyze research aimed at improving overheads in deep learning training and memory consumption. Section 3 introduces the proposed model and suggests a lightweight fire detection mechanism. Section 4 analyzes the experimental environment, content, and results, and Section 5 concludes with an introduction to future research.

2 Related Works

Various studies are being conducted to address issues like training time and memory consumption in data learning using images. This section compares and analyzes previous research, describing the limitations of past studies and the contributions of our proposed research.

In the study proposed by [8], a lightweight, intelligent CNN model was designed to reduce the computational cost of the model. The research addressed power consumption limitations when converting analog signals to digital signals and the computational cost aspects of the image sensor module. Two lightweight CNN models were implemented by reducing the bit precision of the analog-digital converter (ADC) to save power and reduce the number of parameters. The paper experimented with the designed pipeline in MobileNetv2 and GhostNet architectures to assess their generalization capability and performance. While the study demonstrated the generalization ability and reduced power consumption of the model, it could not resolve the slight decrease in model accuracy when reducing ADC bit precision. Additionally, there were limitations related to the dataset, this paper uses high-quality, high-capacity, advanced datasets to improve model performance, making it unsuitable for use in lightweight models.In the study [9] aimed at addressing power consumption in image and video processing and computational cost issues of computer vision applications, an intelligent compression system was proposed to solve the power consumption problem during wireless capsule endoscopy video processing. A deep learning-based classification feedback loop was proposed to determine the importance of images. Important images were enhanced to include additional content, while the less important ones were compressed into lower quality for storage. In this study, we conducted compression and classification experiments on wireless capsule endoscopy(WCE) videos to evaluate the performance of the proposed model, verify the gain of the intelligent compression system, and predict the number of additional transmittable images. The experiments demonstrated the contributions of the study by verifying that achieving high compression rates and classification accuracy is possible while maintaining video quality. However, we did not consider the processing time complexity, and the learning and experiments were limited to specific gastrointestinal organs and lesion presence in the data, making it unclear whether we could achieve the same performance in other learning scenarios.

Study [10] aimed at enhancing the speed of predicting anomalies to detect fire situations. It is emphasized that while recognizing patterns with high accuracy is vital, optimization for real-time execution is also critical. The research adopts the capabilities of Deeplabv3+ and the OpenVINO toolkit to propose an approach close to real-time detection, with experiments and evaluations focusing on process acceleration. The results showed an achieved inference process acceleration of 70.46% to 93.46%. When using a GPU with FP16 precision, the inference process speed was approximately double compared to FP32. This study contributes by considering the accuracy of the detection model and process acceleration and speed in time complexity. However, its limitation lies in analyzing only the impact from a temporal perspective without considering memory availability and accuracy.

In a study [11] using a CNN model trained on actual fire incident images, a custom framework for fire detection was presented using transfer learning. The gradient-weighted class activation mapping (Grad-CAM) method was employed to visualize the fire and pinpoint its location. Experiments were conducted using a composite large-scale dataset formed by merging the fire detection dataset, DeepQuestAI, Saied, Carlo, and Bansal datasets, and the detection performance was evaluated. Experimental results revealed that while the detection accuracies of GoogLeNet, VGG16, and ResNet50 were 88.01%, 64.48%, and 92.54%, respectively, the proposed EfficientNetB0 model exhibited an improved accuracy of 92.68%. However, while traditional research analyses considered model lightweightness and computational costs, this study did not further analyze other metrics besides accuracy. Moreover, while the study introduced EfficientNetB0 as a better method, supposedly lighter than the similarly performing ResNet50, it does not provide concrete evidence to confirm the lightweight nature of the model.

Table 1 summarizes the preceding research that was analyzed.

Table 1. Related research summary table.

Ref.	Features	Limitation
[8]	- Research on lightweight, intelligent CNN models for reduced computational cost	- Uses high-quality, high-capacity datasets - It unsuitable for use in light-weight models.

[9]	- Proposed method to reduce power consumption by decreasing the bit precision of ADC	- Failed to address the decrease in model accuracy when ADC bit precision is reduced
	- Research on intelligent compression systems to address power consumption issues during wireless capsule endoscopy video processing	- Time complexity was not considered
	- Proposed deep learning-based classification feedback loop based on importance	- The data used for training and experiments was limited to specific conditions such as lesions and specific digestive organs
[10]	- Acceleration of the process speed for fire situation detection models	- Various complexities are mentioned, but only time complexity is considered, without accounting for spatial complexities like memory availability
	- Research on optimization for real-time execution	- Did not conduct performance analysis
[11]	- Proposed fire detection framework using transfer learning	- While lightweight and computational cost aspects are mentioned, these metrics are not considered in the experiments
	- Fire visualization and location identification using Grad-CAM	- No evidence is provided to support the claim of proposing a lightweight model

This research showed that not many actively considered optimization among traditional image and video processing studies. Most previous studies either analyzed performance aspects alone or focused on optimization excluding performance, thereby conducting performance analyses limited to specific areas. Some studies that considered accuracy and complexity simultaneously couldn't resolve the trade-off relationship where an increase in accuracy led to increased complexity and improving the complexity aspect resulted in a decrease in accuracy. Therefore, we proposed a mechanism that detects fire by finding the optimal resolution point while maintaining the CNN model performance to address the trade-off issue and enhance fire detection efficiency.

3 Proposed Mechanism

This section details the proposed preprocessing steps and mechanism, elaborating on each stage in depth. First, we explained the criteria used to divide the fire dataset used in the experiment into large fires, medium fires, and small fires. We then discuss how adjusting the resolution helps determine two threshold values. Subsequently, based on the details mentioned above, we discussed the proposed lightweight fire detection mechanism.

3.1 Adjusting Resolution

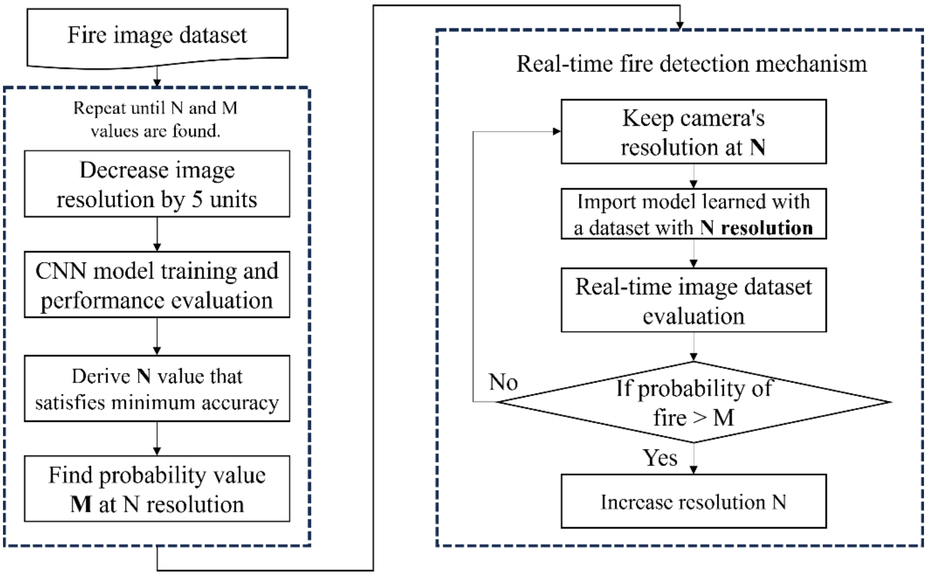

Fig. 1. Flowchart of the mechanism for real-time fire detection.

This study aimed to reduce memory while maintaining performance by reducing image resolution to a point where accuracy is sustained. However, for data where the target object size being detected affects performance, performance differs based on that size. For example, in a medical imaging dataset for tumor detection, one can differentiate between early, middle, and terminal stages based on the tumor size. The early stage would require higher resolution compared to the advanced stage. For reliable experiments, it is necessary to measure the performance separately based on the size of the dataset.

The flow of image resolution adjustment is depicted on the left side of Fig. 1. The original image dataset has a dimension of 224 pixelsAfter adjusting it from 100 to 1, we aimed to identify the resolution point N where performance remained close to the original. When fire is detected using the model with the lowest performance, we calculate the predicted probability estimates for fire classification to derive the average detection likelihood, denoted as M.

3.2 Lightweight fire detection model

After deriving N and M, we proposed a lightweight fire detection mechanism, the diagram shown on the right side of Fig. 1. N represents the threshold value for the minimum resolution, while M serves as the real-time fire detection threshold. In the proposed mechanism, surveillance cameras operate at resolution N, but if they detect a probability exceeding M, they update to a higher resolution. In this context, 'probability' refers to the model's estimation of the likelihood of a fire. When the resolution is 'N,' if the probability exceeds 'M,' the model increases the resolution and performs the detection again in the zone where all models converge in accuracy. If the probability surpasses the threshold 'M,' it is classified as an anomaly.

4 Evaluation

In this section, we describe the experimental environment for implementing and testing the proposed model, mention the content of the experiments, and discuss the results.

4.1 Experimental Environment

We conducted the experiments in an environment with an Intel(R) Core(TM) i9-10850K CPU, 32.0GB RAM, and 930GB Memory, running on the Windows 10 Pro operating system. The tools used were Anaconda3 and Python version 3.10.9. Table 2 provides information on the modules used.

Table 2. Table of used modules.

Module Name	Version
keras	2.10.0
sklearn	1.0.2
numpy	1.23.5
matplotlib	3.5.3
tensorflow	2.10.0
glob	2.69.1
pandas	1.4.2
seaborn	0.11.2

The fire-detection dataset is used [12], an image dataset for detecting fires. This study only used a portion of the dataset, and the fire images were manually verified and categorized into large, medium, and small fires. A large fire is where the fire occupies more than half of the image, a medium fire occupies less than half but more than a quarter of the image, and a small fire takes up less than a quarter of the image. Each large, medium, and small fire is trained separately, and the control group of normal images is used identically in all three models. Table 3 shows the ratio and number of images used in each experiment.

Table 3. Distribution of datasets used by experiment.

Experiment	Image Type	Train	Test	Valid
Large-fire Classification	Large-fire image	140	40	20
	normal image	140	40	20

medium-fire Classification	medium-fire image	140	40	20
	normal image	140	40	20
small-fire Classification	small-fire image	140	40	20
	normal image	140	40	20

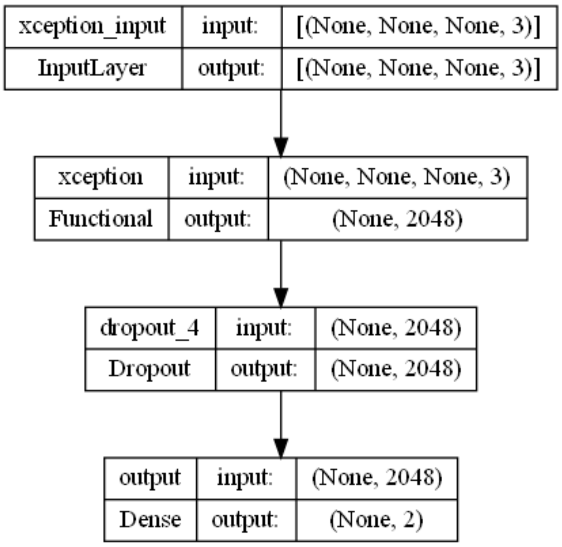

Fig. 2. Layers and input values for the exception model being used.

In the experiment, we used a transfer learning CNN model. Transfer learning models utilize pre-trained models, which can deliver good performance even with data. This made them frequently used models for training with limited images. Fig. 3 shows the operational scenario of the real-time fire detection mechanism. As for other parameter values, we used three channels, the Adam optimizer, and binary_crossentropy, for the loss function. We conducted the training for ten epochs.

4.2 Adjustment of fire image resolution

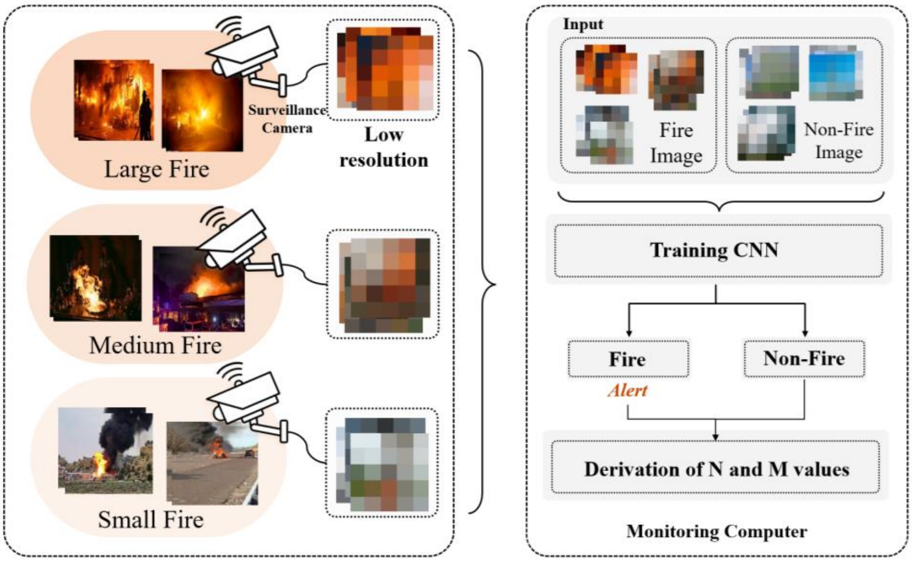

Fig. 3. Operational scenarios of real-time fire detection mechanism

The first experiment aimed to identify the image resolution range where performance is maintained. We reduced the image resolution from 100 to 1 and conducted a binary classification of fire and non-fire, after which we measured the performance. The experiment adjusted the resolution from 100 to 5 pixels in increments of 5. However, since the performance converged from 20 to 100 pixels, we only visualized and analyzed from 5 to 20 pixels. Fig. 4 shows the graph depicting the accuracy according to image resolution.

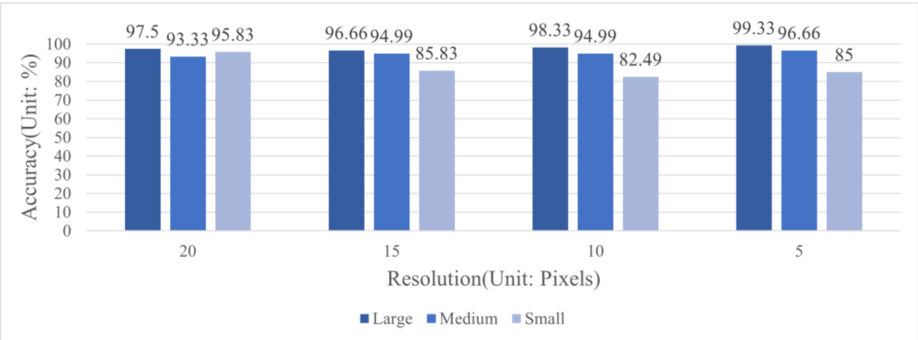

Fig. 4. Evaluation results of detection accuracy by resolution.

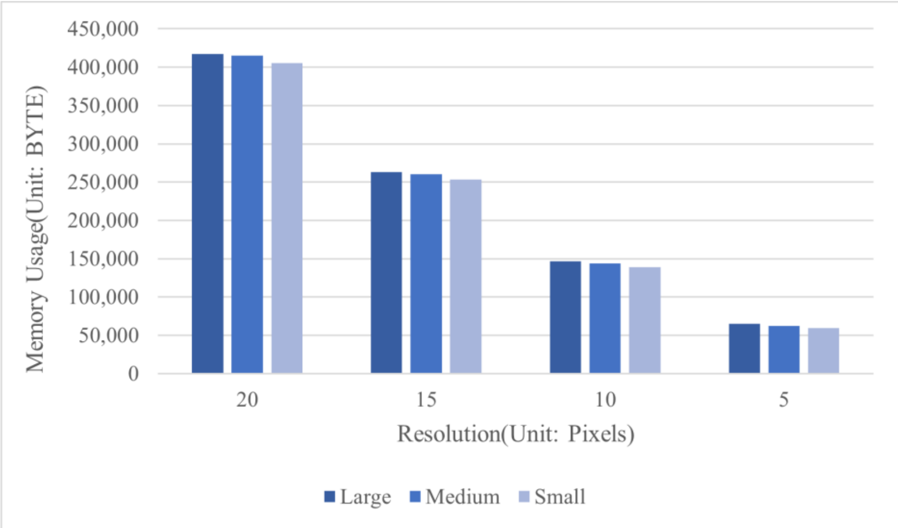

Fig. 5. Evaluation results of memory usage by resolution.

For the large-fire category, the model maintains an accuracy of 99.3 up to a resolution of 5 pixels. The medium-fire maintains a high accuracy of 96.6 at 5 pixels. However, for the small-fire category, even though it sustains a high accuracy of 95 at a resolution of 20 pixels, it drops to a lower performance of 85 when the resolution is at 5 pixels. Therefore, for each dataset, the maximum points where the performance is maintained while reducing the image resolution are confirmed to be 5 for both large- and medium-fire and 20 for small-fire. Fig. 5 shows the evaluation results of the memory usage at each resolution. The original size of 224 pixcels consumes approximately 4.6 million bytes. At the performance retention point for large-fire and medium-fire, which is 5 pixcels, it uses 59,550~65,307 Bytes, while the small-fire at a resolution of 20 utilizes 405,496~416,853 Bytes. This indicates that large- and medium-fire can reduce memory size by up to 70 times, whereas small-fire can save memory by a factor of 10.

In the experiment mentioned above, the small-fire detection demonstrated the least effective performance. However, fires typically spread from small to larger ones, and detecting the fire when it is still a small flame is crucial. Therefore, in the subsequent experiment, we will detect fire using the small-fire dataset to devise an efficient and lightweight fire detection algorithm.

4.3 Evaluation of a lightweight fire detection model

The second experiment evaluated a lightweight fire detection model system for enhanced memory efficiency and effective detection. The proposed mechanism increased the resolution when the probability exceeded a certain threshold, up to a maximum of 20 pixels. The proposed model aimed to detect fires when they are small, so the experiment primarily focused on small-fire detection from the three tests previously conducted. Performance is assessed by measuring the probability, representing the likelihood of matching a particular label. We used the predict function provided by scikit-learn for this purpose.

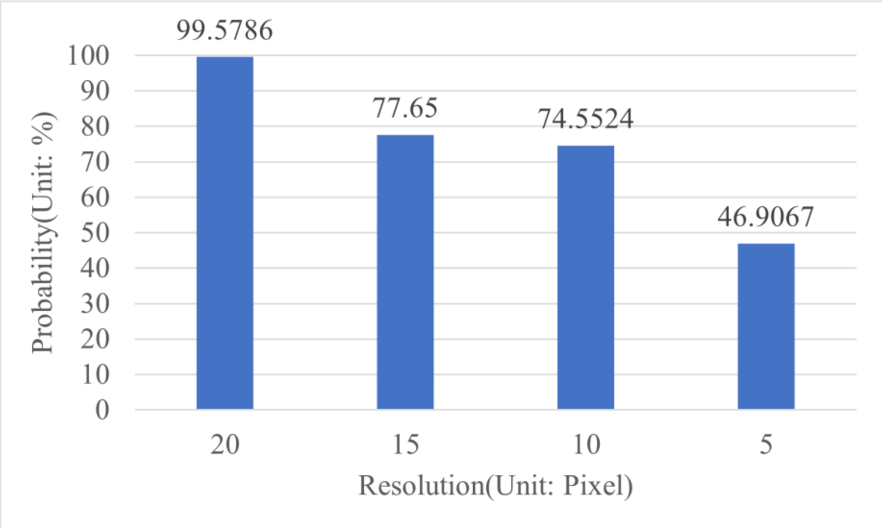

Fig. 6. Evaluation results of probability by resolution.

The probability based on the resolution for small fire is represented in Fig. 6. When the resolution was at 5, it displayed a probability of 46% for fire data. At 10, it showed 74%, and at 20, it converged to 99%.

5 Conclusion

In this study, we proposed a lightweight fire detection model to address the conventional deep learning research limitation of balancing performance with cost. We adjusted the resolution of the images and evaluated the performance for each resolution to determine the threshold value of the proposed model. For the large-fire and medium-fire datasets, a 99.3% accuracy was demonstrated at a resolution of 5, proving 70 times more memory efficient than the original. Furthermore, the small-fire dataset exhibited a 95% accuracy at a resolution of 20, demonstrating it to be ten times more memory efficient. Subsequently, we proposed a two-stage fire detection mechanism, focusing on the small-fire dataset with the lowest performance. This proposed mechanism adjusted the resolution based on the probability of deemed fire and used the measured probability from the small-fire dataset as its threshold. Ultimately, the proposed model proved to be approximately 31 times more memory efficient while maintaining fire detection performance.

However, this study utilized a limited dataset, and various variables may have influenced the experimental results. To derive more reliable results, repetitive testing with vast data is necessary. Therefore, in the future, we plan to conduct experiments targeting a broader and more diverse dataset and aim to derive trustworthy outcomes through repeated experiments.

Acknowledgements

This work is partly supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government(MOTIE) (P0008703, The Competency Development Program for Industry Specialist), and the MSIT under the ICAN (ICT Challenge and Advanced Network of HRD) program (No. IITP-2022-RS-2022-00156310) supervised by the Institute of Information & Communication Technology Planning & Evaluation (IITP).

References

1. Soheil, V., Kin-Choong, Y.: A CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras. applied sciences 12(3) (2022)
2. Trong, N., Jean, M.: Anomaly Detection in Video Sequence With Appearance-Motion Correspondence. In: 2019 IEEE/CVF International Conference on Computer Vision, ICCV, Seoul Korea (2019)
3. Asif, K., Quadri S.M.K., Saba, B., Junaid, L.: Deep diagnosis: A real-time apple leaf disease detection system based on deep learning. Computers and Electronics in Agriculture, vol. 298 (2022)
4. Amreen, A., Sweta, J., Mahesh, G., Swetha, V.: Tomato plant disease detection using transfer learning with C-GAN synthetic images. Computers and Electronics in Agriculture, Vol. 187 (2021)
5. Waqas, A., Umair, N., Muhammad, H. Y., Sergio, A. V.: Lightweight CNN and GRU Network for Real-Time Action Recognition. In: 2022 12th International Conference on Pattern Recognition Systems, ICPRS, Saint-Etienne, France (2022)
6. Daiki, A., Tomio, G.: Improving Image Quality Using Noise Removal Based on Learning Method for Surveillance Camera Images. In: 2022 IEEE 4th Global Conference on Life Sciences and Technologies, LifeTech, Osaka, Japan (2022)
7. Muhammad, A. K., Ridha, H., Aiman, E., Moncef, G.: Distributed Inference in Resource-Constrained IoT for Real-Time Video Surveillance. IEEE Systems Journal 17(1), 1512 – 1523 (2022)
8. Wilfred, K., Amandeep, K., Deepak, M.: An Intelligent System with Reduced Readout Power and Lightweight CNN for Vision Applications. IEEE Transactions on Circuits and Systems for Video Technology (2023)
9. Dallel, B., Said, B., Mohammed, K.: An intelligent compression system for wireless capsule endoscopy images. Biomedical Signal Processing and Control. Vol. 70 (2021)
10. Medina, K., Adnan, S., Amila, A., Emir, B.: Early Stage Flame Segmentation with Deep Learning and Intel's OpenVINO Toolkit. In: 2023 XXIX International Conference on Information, Communication and Automation Technologies, ICAT (2023)
11. Saima, M., Fayadh, A., Sarfaraz, M., Musheer, A., Emine, S. G., Kemal, P.: Attention based CNN model for fire detection and localization in real-world images. Expert Systems with Applications, Vol. 189 (2022)

12. kaggle, https://www.kaggle.com/datasets/tharakan684/urecamain/data, last accessed 2023/10/06
13. So-Eun, J., Ye-Sol, O., Ye-Seul, K., Yeon-Ji, L., Il-Gu, L.: Two-Step Feature Selection Technique for Secure and Lightweight Internet of Things. 2023 32nd International Conference on Computer Communications and Networks, ICCCN, Honolulu, HI, USA (2023)
14. Sun-Woo, Y., Na-Eun, P., Il-Gu, L.: Wake-Up Security: Effective Security Improvement Mechanism for Low Power Internet of Things. Intelligent Automation & Soft Computing, 37(3), 2897-2917 (2023)