Vision-based Early Fire and Smoke Detection for Smart Factory Applications Using FFS-YOLO

Duc Tri Phan[1], Kim-Hui Yap[1], Kratika Garg[2], Boon Siew Han[2]

***Abstract*— Early-stage fire and smoke detection through visual analysis is crucial for industrial safety and hazard prevention. However, detecting fire and smoke in factories using surveillance cameras poses challenges due to the small size of target objects. To address these challenges, we introduce a refined single-stage detector called FFS-YOLO (Factory Fire Smoke – YOLO). Our approach incorporates the Parameter-Free Attention Module (SimAM) and ResNet-SimMix module into the Backbone and Head of YOLOv7 to enhance key feature extraction. Additionally, we modify the model architecture by adding an extra prediction head to facilitate the fusion of features at multiple scales, specifically for small-scale object detection. Experimental results conducted on our fire and smoke dataset demonstrate the effectiveness of the FFS-YOLO model, achieving an average mAP, Precision, and Recall of 0.92, 0.91, and 0.90, respectively. The performance of the proposed model outperforms existing relevant competitors in the field. The findings of this research contribute significantly to the advancement of early fire detection and prevention in factory settings.**

***Index Terms*— Fire and Smoke Detection, YOLO, Parameter-Free Attention Module, Small Object Detection, Industrial Safety.**

I. INTRODUCTION

THE occurrence of unrestrained fires in factories poses a severe threat to infrastructure, and human lives [1]. Detecting fire and smoke in their early stages is a pressing concern for industrial safety and decision-making management [2]. However, current approaches for early detection of fire and smoke often rely on manual observation or sensors [1]. Manual observation-based methods have their limitations, particularly in the presence of intermittent interruptions or distractions, making them impractical for continuous and reliable monitoring [3]. On the other hand, traditional sensor-based systems rely on smoke particle sampling or relative heat to detect fire, but they often suffer from significant time delays and require strong fires or close proximity to sense the fire [3]. Additionally, they may not be able to comprehensively cover detection areas, especially in wide-open spaces, due to environmental variations [3]. In general, current fire and smoke detection techniques struggle to meet the demands of modern industrial processes and safety requirements [1-3].

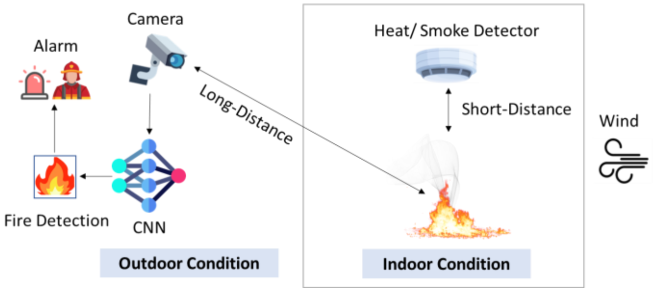

Fig. 1. The schematic diagram of the vision-based fire and smoke detection system for industrial safety.

In recent years, image-based fire and smoke detection methods leveraging visual processing and deep neural network techniques have emerged as promising solutions to address the challenges of fire detection [1]. Geetha *et al.* conducted a comprehensive survey on the use of image processing (IP), machine learning (ML), and deep learning (DL) techniques for fire and smoke detection, covering various aspects such as datasets, methodologies, challenges, and future work in this field [4]. Meanwhile, Khan *et al.* published a review article that focused on video analysis for flame and smoke detection using both traditional ML algorithms and DL algorithms [5]. Chaturvedi *et al.* conducted a chronological review of fire detection systems and approaches, covering both traditional and advanced techniques [1]. With the rise of DL technology and the expansion of fire and smoke datasets, numerous deep learning-based techniques have been developed for fire and smoke detection, as described by Sathishkumar *et al.* [6], Wang *et al.* [7], Mohammad *et al.* [8], and Dewangan *et al.* [9], among others. Although IP and DL algorithms have made significant achievements in smoke and fire detection, early-stage detection remains an area with limited research. To address this issue, this study proposes a novel single-stage detector called FFS-YOLO for early-stage fire and smoke detection in surveillance videos.

The primary contributions are summarized as follows:

This research is supported by the Agency for Science, Technology and Research (A*STAR) under its IAF-ICP Programme I2001E0067 and the Schaeffler Hub for Advanced Research at NTU.

[1] Duc Tri Phan and Kim-Hui Yap are with School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (Email: ductri.phan@ntu.edu.sg, ekhyap@ntu.edu.sg).

[2] Kratika Garg and Boon Siew Han are with Schaeffler Hub for Advanced Research, Nanyang Technological University, Singapore
(Email: gargkat@schaeffler.com, hanbon@schaeffler.com)

2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP) | 979-8-3503-3893-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/MMSP59012.2023.10337640

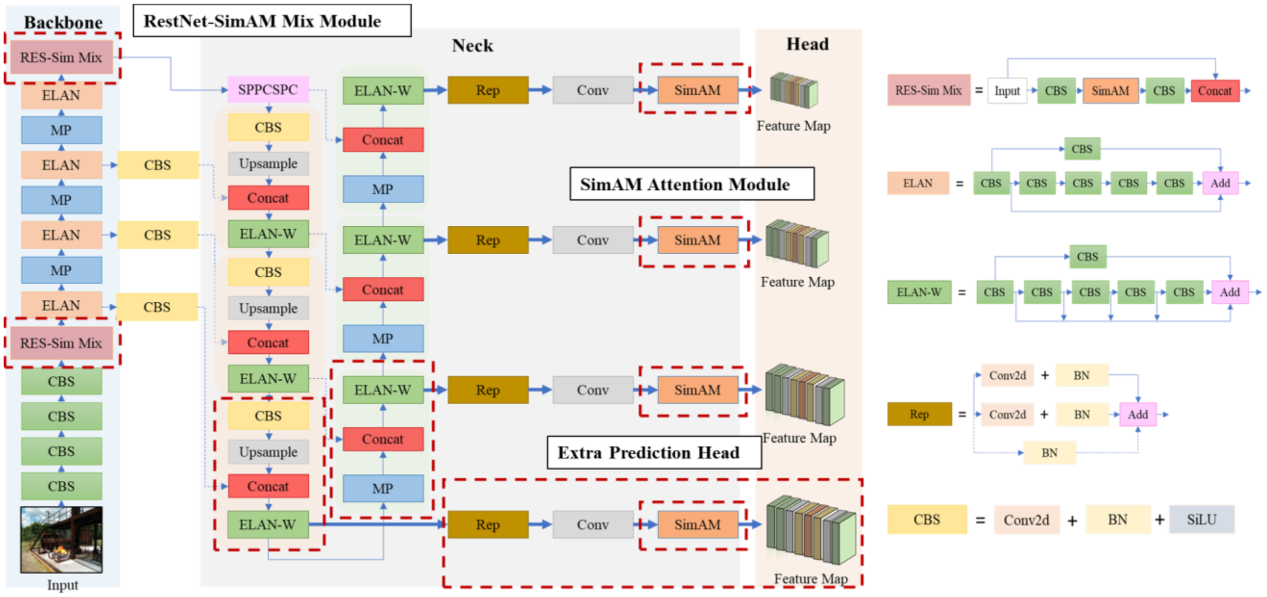

Fig. 2. The architecture of the FFS-YOLO for early fire and smoke detection.

- The YOLOv7 network is enhanced with the SimAM attention module to improve its capability to identify regions of interest. This module enables the network to retain important features of amplification, reducing information loss and facilitating better learning of useful information.
- To increase the performance of the proposed model in detecting small objects, the network structure of the original YOLOv7 is modified based on the concept of tiny object detection. An additional prediction head is added to enhance feature extraction and enable more accurate detection of small objects.
- The proposed FFS-YOLO model is evaluated on our factory fire and smoke dataset, where it outperforms other one-stage detector networks in detecting small fires. This makes the FFS-YOLO model suitable for real-time deployment on surveillance cameras placed at long distances, which can significantly enhance fire prevention for industrial safety.

II. Related Works

Fire and smoke recognition have been widely investigated using convolutional neural networks (CNNs) such as ResNet [10] and KutralNet [11]. However, recent works have explored the utilization of lightweight models specifically designed for the detection task. For instance, Otabek *et al.* introduced Light-FireNet, a lightweight CNN that achieved a test accuracy of 97.83% on their dataset [12]. Ayala *et al.* proposed another lightweight model, which combined the Octave convolution with ResNet, obtaining an average validation accuracy of 87.44% on four benchmark datasets for fire and smoke detection [13]. Seyd *et al.* compared their Fire-Net model against MSR-U-NET and other machine learning methods [14]. Additionally, Ali *et al.* employed the highly efficient CNN architecture EfficientNet-BO for fire emergency recognition from images [15]. Moreover, Yuan *et al.* presented a novel algorithm for smoke image classification that utilizes high-order local ternary patterns (LTP) with local preservation projection, which outperformed existing methods [16]. Furthermore, Zhang *et al.* introduced the use of a depthwise separable convolution (IDCNN) in sequential CNNs for image-based smoke detection, resulting in improved model efficiency [17].

III. Method

A. Overall Architecture of FFS-YOLO

FFS-YOLO, which is based on YOLOv7 [18], is composed of three main modules: the backbone, neck, and head, as shown in Figure 2. The backbone module of FFS-YOLO comprises three primary components: CBS, RestNet-SimAM Mix Module, extended efficient layer aggregation network (E-ELAN), and MaxPool (MP). These components are employed to extract valuable information from the input module. In the neck module, the Feature Pyramid Network (FPN) architecture is combined with CBS, Convolutional Spatial Pyramid Pooling (Sppcspc), E-ELAN, and MP to enhance feature extraction. The head module employs a REP structure to adjust the number of feature outputs from the neck module to reduce network complexity while maintaining high predictive accuracy.

However, the original YOLOv7 network is not optimized for small object detection, which limits its effectiveness in early-stage fire and smoke detection [19]. To overcome this limitation, we propose modifications to the network structure and integrate attention modules to enhance the performance and accuracy of the proposed model in detecting small objects. Firstly, we integrate the RES-Sim Mix module into the backbone module to extract key information more effectively. Secondly, we introduce additional prediction heads in both the neck and head modules, inspired by the concept of tiny object detection modules, to enhance the capability of algorithm in detecting small objects. Moreover, the SimAM attention module is incorporated into the FFS-YOLO network for better learning of useful information. The SimAM attention module highlights important target features in the shallow network while suppressing other irrelevant features, leading to improved detection performance for the small target [20].

B. Extra Prediction Head

The detection of small objects located at a long distance in the fire poses a challenge and has a negative impact on the early fire alarm system. For address the issue, we included an up-sampling module in the neck layer to increase the image resolution and generate more informative feature maps. In comparison with the three detection heads, the four detection heads achieve more multi-scale detection. This enhancement improves detection stability and mitigates the adverse effects of significant scale variations in objects. As shown in Figure 2, the additional up-sampling module and prediction head in the neck and head layers enhance the fusion of features at multiple scales and improve the robustness of detection at different scales, despite the associated increase in computational and memory costs.

B. SimAM Attention Module

Attention modules are widely employed in neural network models to enhance feature extraction and highlight the salient objects of interest. However, almost attention modules allocate weights along the channel dimension, resulting in increased model parameters and complexity. In contrast, SimAM is an attention mechanism that directly assigns 3D attention weights to feature maps without requiring additional parameters [20]. This approach improves the feature extraction capability while maintaining a lightweight and efficient design. SimAM identifies important neurons and assigns them higher priority based on linear separability from other neurons in the same channel. The minimum energy function of each neuron is determined as follows:

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t-\hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{1}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are mean and variance, and can be calculated as:

$$\hat{\mu} = \frac{1}{M}\sum_{i=1} x_i \tag{2}$$

$$\hat{\sigma}^2 = \frac{1}{M}\sum_{i=1}^{M}(x_i - \hat{\mu})^2 \tag{3}$$

t and x_i are the target neuron and other neurons in the channel, λ is the hyper-parameter and M is the is the total count of neurons on a one channel.

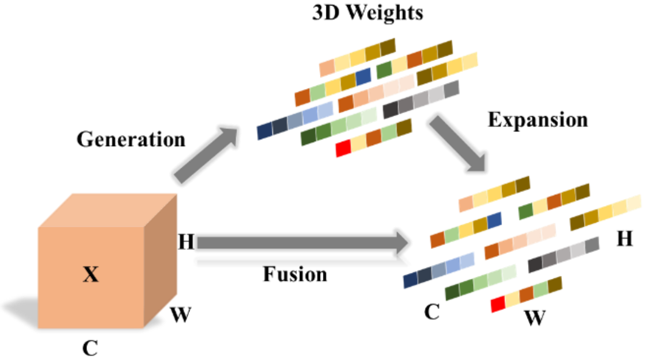

Fig. 3. Sim AM with full 3D weights for attention.

The energy function value in equation (1) determines the distinctiveness of a neuron compared to its neighbouring neurons, with smaller values indicating greater linear separability. Specifically, the importance of a neuron can be calculated as 1 divided by the energy function value. Research conducted by Lingxiao Yang *et al.* has demonstrated that integrating SimAM into various classification and detection models enhances feature extraction performance without adding extra model parameters [20].

D. RestNet-SimAM Mix Module

The incorporation of the ResNet-SimAMmix module in the Backbone component of FFS-YOLO ensures the consistency of the extracted feature information. This module adopts the bottleneck structure of ResNet but substitutes the 3x3 convolution with the SimAMmix module, as illustrated in Figure 4. This ResNet-SimAMmix module allows for adaptive focus on different regions and the capture of more informative features. As a result, the ResNet-SimAMmix module enables the network to reach deeper depths without encountering gradient disappearance.

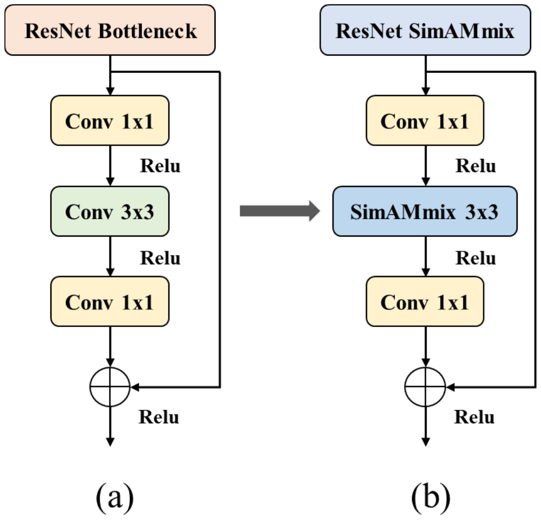

Fig. 4. The structure diagram of (a) ResNet and (b) ResNet-SimAMmix.

IV. Experiment and Analysis

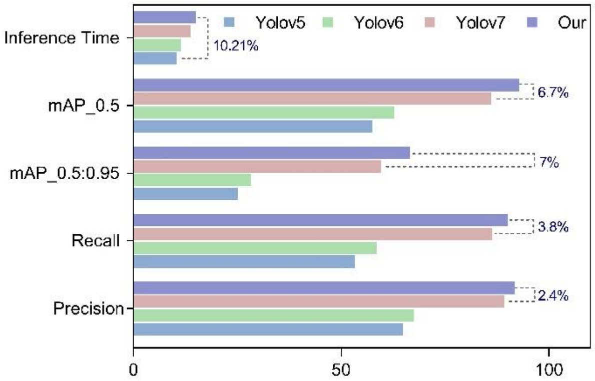

Fig. 5. Comparison of our model with other networks.

A. Datasets and Experimental Settings

The fire and smoke dataset, comprising 8000 images obtained from various sources such as Kaggle, Roboflow, NIST, SKLFS, and the Internet, is specifically designed for early fire and smoke detection in factory scenes. To mitigate the issue of overfitting, data augmentation techniques such as angle rotation, saturation adjustment, and image flipping were applied. The fire and smoke dataset was randomly divided into training, validation, and test sets, following an 8:1:1 ratio.

For training FFS-YOLO, we utilized PyTorch 1.13.1 and performed inference on NVIDIA RTX A5000 GPUs. To accelerate the training process, a partially pre-trained YOLOv7 model was employed. The proposed model was trained on the FFS dataset for 200 epochs using the SGD optimizer with an initial learning rate of 0.1. The input image size was set to 640 x 640 pixels, and a batch size of 24 was used. Additionally, we implemented the evolve hyper-parameters method to continuously optimize the hyper-parameters throughout the training process.

B. Evaluation Metrics

The primary evaluation metrics for the model in this research include precision (P), recall (R), and mean average precision (mAP). Target detection is determined based on a threshold for the intersection over union (IOU) between the prediction frame and the target frame, which is set to a value greater than 0.5.

C. Algorithm Comparison

We evaluated FFS- YOLO and compared its performance with other one-stage detectors, including YOLOv5, YOLOv6, and the original YOLOv7. The results, presented in Table I and Fig. 5, demonstrated that FFS- YOLO outperformed YOLOv7 in terms of precision and recall, with values of 0.91 and 0.9, respectively. These metrics were 2.4% and 3.8% higher than those of YOLOv7, respectively. Additionally, the proposed model achieved higher mAP values at IOU thresholds of 0.5:0.95 and 0.5 (0.665 and 0.928, respectively) compared to YOLOv5, YOLOv6, and YOLOv7. However, FFS-YOLO exhibited a slower inference time (IT) due to its larger size. Despite this drawback, the proposed network achieved improved accuracy and maintained a reasonable balance between detection accuracy and speed, making it well-suited for deployment in complex factory environments for fire and smoke detection.

TABLE I
COMPARISON OF THE FFS-YOLOV 7 WITH OTHER NETWORKS.

Model	Input	P	R	mAP_0.5	IT
YOLOv5	640^2	0.64	0.53	0.57	10.4ms
YOLOv6	640^2	0.67	0.58	0.62	11.5ms
YOLOv7	640^2	0.89	0.86	0.83	13.7ms
Our	640^2	0.91	0.90	0.92	15.1ms

The detection results are visualized in Figure 6, which demonstrates the successful identification of small fires and smoke in a factory with significantly higher recognition accuracy. Our method shows minimal missed detections and false positives, indicating its effectiveness in accurately detecting early fire and smoke in factories.

D. Ablation Study

To evaluate the effectiveness of our proposed methods in improving detection tasks, we performed ablation tests by progressively incorporating optimized modules from each layer. The detailed results of these tests are presented and compared in Table II.

Effect of extra prediction head. The addition of an extra detection head in the original YOLOv7 model results in an increase in the number of parameters and network layers from 415 to 509. Moreover, GFLOPS (Giga-Floating Point Operations per Second) increases from 105.1 G to 201.7G. However, it also enhances the sensitivity and effectiveness of the detector in detecting small foreground objects. The results presented in Table II confirm that the mAP_0.5 of the extended YOLOv7 model increases from 0.861 to 0.912, demonstrating the efficacy of incorporating the extra detection head.

Effect of SimAM module. Despite incorporating the attention mechanism in both the backbone and head, the size and parameters of the networks remained largely unchanged. The SimAM attention module enables the network to focus on important regions while still maintaining a similar model size and parameter count. This module allows for the flexible assignment of 3D attention weights to feature maps, thereby enhancing the network feature processing capabilities.

TABLE II
ABLATION TESTS AFTER OPTIMIZATION OF YOLOV7.

Model	Inference Time	mAP	Layers	Parameters	GFLOPS
YOLOv7	13.7ms	0.86	415	37.2M	105.1G
YOLOv7+SimAM	14.1ms	0.88	443	35.3M	100.8G
YOLOv7+ Extra Prediction Head	14.8ms	0.91	509	79.8M	201.7G
Our	15.1ms	0.92	537	80.9M	203.4G

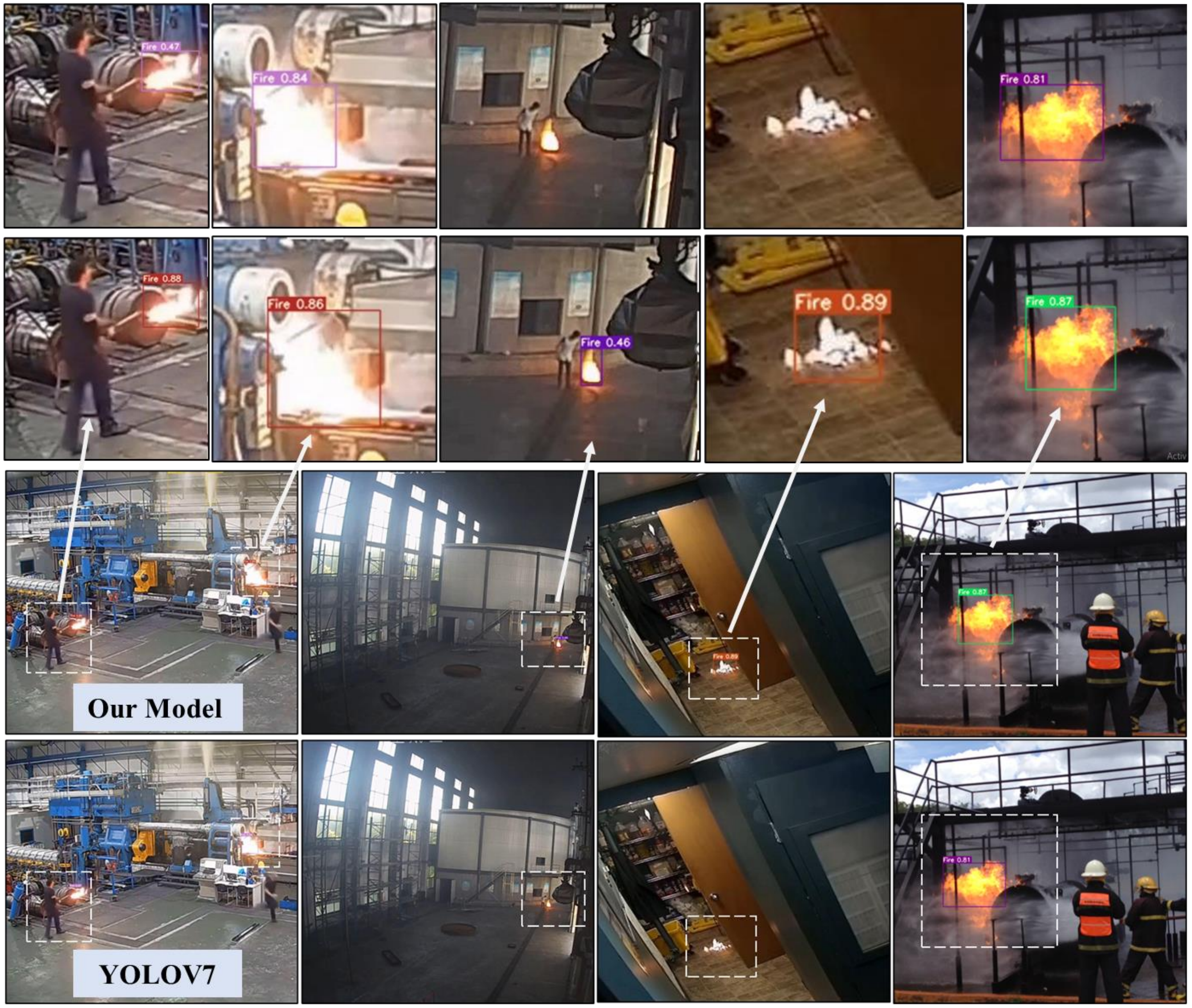

Fig. 6. Fire and smoke detection results of Yolov7 and our proposed model

Effect of the model ensemble. The FFS-YOLO model combines the optimization capabilities of the extra prediction head and the SimAM module, resulting in improved object detection ability of the final model. Comparing the mAP indexes reveals that the proposed method outperforms other models that have separate extra prediction heads and SimAM modules. While these additions of attention module, and extra prediction head can significantly improve the model's performance and ability to capture complex patterns, they also come with an increase in computational and memory requirements, making it necessary to strike a balance between model complexity and available resources.

V. Conclusion

The proposed FFS-YOLO model in this study has shown superior performance compared to existing single-stage detection models. The integration of the SimAM module into the network has enhanced the feature extraction ability for fire detection. The experimental results demonstrate significant improvements, with the proposed method achieving a 6.7% increase in mAP_05, as well as 2.4% and 3.8% improvements in precision and recall, respectively, compared to the original YOLOv7 network. The FFS-YOLO network exhibits high detection performance and robustness for small object detection, making it a promising solution for real-time fire detection in complex factory environments. The findings of this study can contribute to improving safety and preventing fire incidents in factories through long-distance surveillance.

Acknowledgment

This research is supported by the Agency for Science, Technology and Research (A*STAR) under its IAF-ICP Programme I2001E0067 and the Schaeffler Hub for Advanced Research at NTU.

References

[1] Chaturvedi, S., et al., *A survey on vision-based outdoor smoke detection techniques for environmental safety.* ISPRS Journal of Photogrammetry and Remote Sensing, 2022. 185: p. 158-187.

[2] Savitha, N. and S. Malathi. *A survey on fire safety measures for industry safety using IOT*. in *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*. 2018. IEEE.

[3] Barmpoutis, P., et al., *A review on early forest fire detection systems using optical remote sensing.* 2020. 20(22): p. 6442.

[4] Geetha, S., C. Abhishek, and C.J.F.t. Akshayanat, *Machine vision based fire detection techniques: a survey.* 2021. 57: p. 591-623.

[5] Khan, F., et al., *Recent advances in sensors for fire detection.* 2022. 22(9): p. 3310.

[6] Sathishkumar, V.E., et al., *Forest fire and smoke detection using deep learning-based learning without forgetting.* 2023. 19(1): p. 1-17.

[7] Wang, M., et al., *FASDD: An Open-access 100,000-level Flame and Smoke Detection Dataset for Deep Learning in Fire Detection.* 2023: p. 1-26.

[8] K Mohammed, R.J.I.J.o.N.A. and Applications, *A real-time forest fire and smoke detection system using deep learning.* 2022. 13(1): p. 2053-2063.

[9] Dewangan, A., et al., *FIgLib & SmokeyNet: Dataset and deep learning model for real-time wildland fire smoke detection.* 2022. 14(4): p. 1007.

[10] Jabnouni, H., et al. *ResNet-50 based fire and smoke images classification*. in *2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. 2022. IEEE.

[11] Ayala, A., et al. *KutralNet: A portable deep learning model for fire recognition*. in *2020 International Joint Conference on Neural Networks (IJCNN)*. 2020. IEEE.

[12] Khudayberdiev, O., et al., *Light-FireNet: an efficient lightweight network for fire detection in diverse environments.* 2022. 81(17): p. 24553-24572.

[13] Ayala, A., et al. *Lightweight and efficient octave convolutional neural network for fire recognition*. in *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. 2019. IEEE.

[14] Seydi, S.T., et al., *Fire-Net: A deep learning framework for active forest fire detection.* 2022. 2022: p. 1-14.

[15] Ali, A.-e.A., et al., *Efficient Net: A Deep Learning Framework for Active Fire and Smoke Detection.* 2023. **3**(02): p. 1-10.

[16] Yuan, F., et al., *High-order local ternary patterns with locality preserving projection for smoke detection and image classification.* 2016. 372: p. 225-240.

[17] Zhang, J., et al., *Compressed dual-channel neural network with application to image-based smoke detection.* 2022. **16**(4): p. 1036-1043.

[18] Wang, C.-Y., A. Bochkovskiy, and H.-Y.M.J.a.p.a. Liao, *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors.* 2022.

[19] Liu, S., et al., *CEAM-YOLOv7: Improved YOLOv7 Based on Channel Expansion and Attention Mechanism for Driver Distraction Behavior Detection.* 2022. 10: p. 129116-129124.

[20] Yang, L., et al. *Simam: A simple, parameter-free attention module for convolutional neural networks*. in *International conference on machine learning*. 2021. PMLR.