2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

GEMN: A Novel Forest Fire Detection Network

Yong Liu[1], Shaochen Jiang[1*] and Yongming Li[1]

[1]College of Computer Science and Technology, Xinjiang University
Xinjiang Multimodal Intelligent Processing and Information Security Engineering Technology Research Center
∗This work was supported by the Xinjiang Uygur Autonomous Region Tianshan Talent Program Project under Grant (2022TSYCLJ0036)

107552403819@stu.xju.edu.cn, scj@xju.edu.cn, lym@xju.edu.cn

Abstract. Forest fires have had a significant impact on global ecosystems and human societies, necessitating the development of efficient and accurate early smoke detection technologies to combat fires. However, current smoke detection technologies face multiple challenges in real- time applications, including issues such as large parameter sizes, high computational complexity, and low detection accuracy in complex scenarios. Therefore, based on YOLOv10, we propose a lightweight, high-precision, real-time smoke detection network called GEMN. Firstly, to reduce the extraction of redundant features, we innovatively designed a GGCA attention mechanism. This mechanism significantly enhances the comprehensiveness of the model's feature extraction by strengthening the representation of important features. Secondly, to lower the computational complexity and parameter count of the model, we introduced a lightweight detection head named EISDH. Thirdly, we incorporated the MPDIoU function. This not only enhances the model's robustness but also simplifies the process of extracting unnecessary features from forest fire targets, further reducing the parameter count in the network model. GEMN demonstrates exceptional performance across three testing benchmark datasets. Notably, on the FFES dataset, compared to the baseline model, our GEMN model achieves a remarkable 0.991% mAP improvement of 2.5%, reaching an accuracy of 96.8%, an increase of 3.7%. Meanwhile, it compresses the parameters to 4.0MB and shortens the inference time to 1.0 milliseconds, showing an approximately 30% improvement over the original model.

Keywords: Image Processing, YOLOv10, Forest Fires

1 Introduction

Smoke detection technology is widely utilized in forest protection and daily life, serving as one of the primary tasks in computer vision methodologies. Early smoke detection approaches predominantly relied on sensors and traditional machine learning algorithms; however, these methods often faced limitations such as deployment difficulties or low accuracy. In recent years, due to its remarkable performance in terms of accuracy and efficiency, deep learning technology has emerged as a new force in the field of real-time object detection, with an increasing number of studies applying it to smoke detection. Nevertheless, existing models encounter challenges in real-time applications: traditional algorithms, owing to their large parameter sizes and high computational complexity, are difficult to deploy on devices with limited resources. These devices possess constrained computing capabilities and are unable to effectively support algorithms with high computational demands, thereby affecting the real-time nature and flexibility of smoke detection. Additionally, although there exist some lightweight smoke detection algorithms, this often comes at the expense of sacrificing detection details and accuracy. In complex scenarios, particularly when detecting thin or small-scale smoke, these algorithms often fail to capture and recognize subtle features, consequently impacting detection accuracy.

To address the aforementioned challenges, we propose a lightweight smoke detection network based on YOLOv10, termed GEMN. This network not only enhances the accuracy and robustness of smoke detection but also reduces the computational complexity and parameter size of the model. The main contributions of this paper include:

To reduce the extraction of redundant features, we innovatively designed a GGCA (Global Guided Channel Attention) attention mechanism. This mechanism significantly enhances the comprehensiveness of feature extraction in the model by strengthening the representation of important features.

To reduce parameters, we have designed a novel detection head named EISDH (Efficient Information Sharing Detection Head). This detection head achieves a reduction in both parameters and computational load while ensuring detection accuracy.

To address the issue of bounding box distortions caused by significant sample variability and to enhance the robustness of the model during the detection process, we have introduced the MPDIoU function. This not only strengthens the model's robustness but also simplifies the process of extracting unnecessary features from forest fire targets, further reducing the number of parameters in the network model.

On the FFES fire smoke dataset, GEMN achieved a mAP50 (mean Average Precision with a 50% Intersection over Union threshold) of 99.1%, with an accuracy rate reaching 96.8%. The number of parameters was reduced to 4.0M, while the detection speed decreased by approximately 30%.

2 Related Work

In the early stages of fire smoke detection, sensors play a crucial role. However, deploying a sufficient number of sensors across vast forested areas poses certain challenges. Additionally, delays in data transmission may affect monitoring efficiency. Traditional machine learning algorithms detect smoke features by combining image processing techniques with machine learning. Research indicates that texture is the most stable and distinctive attribute among smoke features [1]. Therefore, many studies focus on how to extract texture features more effectively to enhance the accuracy of smoke detection. Based on these features, various smoke detection algorithms have been developed. For instance, Almgiretal. [2]conducted color-based clustering analysis on smoke images to identify smoke regions.

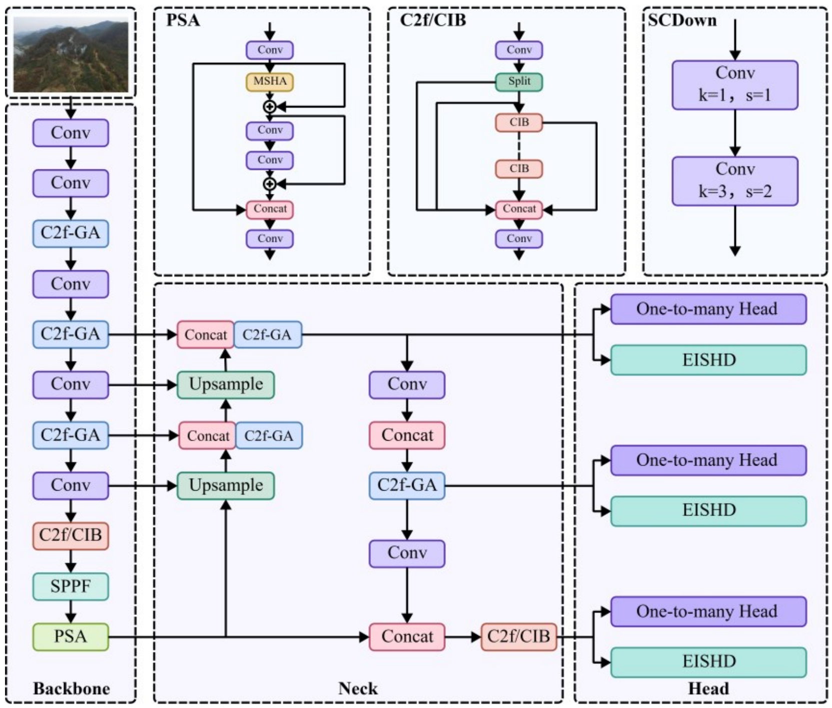

Fig. 1. GEMN Overall Architecture

Within these smoke regions, texture features of Local Binary Patterns (LBP) [3] are extracted, and a coexistence model with color features is established to recognize smoke objects. Liu et al. [4] defined a contour coefficient variable for a local binary model, which is used for segmenting smoke images. Although traditional machine learning methods have achieved some success in smoke and fire detection, they rely on extensive feature extraction, typically requiring significant human and time resources.

Moreover, due to the significant variability of smoke and flame features, manually extracted features struggle to cover all possible scenarios, thereby limiting the generalizability of these methods across different environmental backgrounds. consequently, methods based on traditional feature extraction face challenges in improving the accuracy of fire smoke detection.

Given the high precision and efficiency demonstrated by deep learning methods in early fire smoke detection, smoke detection approaches based on deep learning have been widely adopted. Zeng et al. [5] improved the accuracy of flame and smoke detection by enhancing Convolutional Neural Network (CNN) models, showing better performance in open spaces. Yang et al. [6], drawing inspiration from MobileNet [7], developed a lightweight CNN fire detection model to increase efficiency and reduce resource consumption. These methods typically employ image classification techniques to detect smoke. However, to precisely locate the position of smoke within images, target detection techniques are necessary. Target detection is divided into single-stage and two-stage methods. Two-stage target detection is represented by the Region Convolutional Neural Network (RCNN) series of algorithms, such as Faster RCNN [8]. Wang et al. [9] conducted smoke detection experiments based on the Faster RCNN network and achieved significant progress. Although two-stage algorithms possess high detection accuracy, their complex calculations and large number of parameters limit their use in real-time detection on mobile or edge devices. In contrast, single-stage target detection algorithms like the YOLO series, with their simpler structures and faster processing speeds, are more suitable for use on these devices. Li [10] proposed an improved version of YOLOv3, named SRNYOLO, which enhanced detection precision and simplified the network structure. Wang et al. [11] incorporated a lightweight MobileNet backbone into the YOLOv4 framework, enhancing the model's robustness by using fire-like images as negative samples. Xue et al. [12] optimized the bounding box loss function of YOLOv5, adopting the SIoU loss function and integrating directional features to accelerate model training and inference. Xue et al. [13] also introduced Light-YOLOv5, a lightweight fire detection model that utilizes the SepViT module and a lightweight bidirectional feature pyramid network, enhancing feature extraction and improving detection accuracy through a global attention mechanism. Although YOLOv5 supports efficient processing of images and videos, there is still room for improvement in detection accuracy.

To address the aforementioned challenges in the field of fire smoke detection, this study innovatively proposes a precise fire smoke detection and recognition solution based on the YOLOv10 deep learning model. As the latest version of the YOLO series, YOLOv10 demonstrates significant performance improvements in smoke detection tasks with its superior recognition accuracy and unprecedented detection speed. This study conducted in-depth optimization of the YOLOv10 model, aiming to further enhance the accuracy and processing speed of fire smoke detection and recognition.

3 Materials and Methods

3.1 GEMN Network Structure

The YOLOv10 network architecture is primarily composed of four key components: the Input layer, Backbone network, Neck network, and Detect Head [14]. The Input layer enhances the data through scaling and normalization operations. The Backbone and Neck networks adopt the C2f structure and the Path Aggregation Network (PAN) structure to accelerate network convergence and perform feature extraction. The Detect Head employs a dual-head design, namely the One-to-many Head and the One-to-one Head, for object localization and classification. The overall architecture of our proposed Global Enhanced Multi-scale Network (GEMN) is illustrated in Figure 1. Built upon YOLOv10, GEMN incorporates our proposed C2f-GA module and Enhanced Interactive Spatial and Depthwise Hierarchical (EISDH) module. These modules reduce redundant computations, more effectively capture fire smoke features, and improve the model's accuracy. Additionally, we introduce the Multi-scale Predicted Intersection over Union (MPDIoU) function, which not only enhances the model's robustness but also simplifies the process of extracting unnecessary features from fire smoke targets, further reducing the number of parameters in the network model.

3.2 C2f-GA

To address the challenge of difficult recognition in complex forest fire scenarios, enhancing the neck network of the model is imperative. Due to varying weather conditions and the movement of image acquisition equipment, geometric distortions can occur in the captured targets, which negatively impact detection performance. By integrating our innovatively proposed Global Guidance Channel Attention (GGCA, with this reasonable supplementation of the potential missing abbreviation based on common naming conventions for technical terms; if the original text has a specific full name, it should be followed) module, this network not only strengthens the representation of key features in complex forest fire scenarios but also improves adaptability to dynamically changing scenes, while simultaneously enhancing detection accuracy.

The architecture of this network is illustrated in Figure 2.

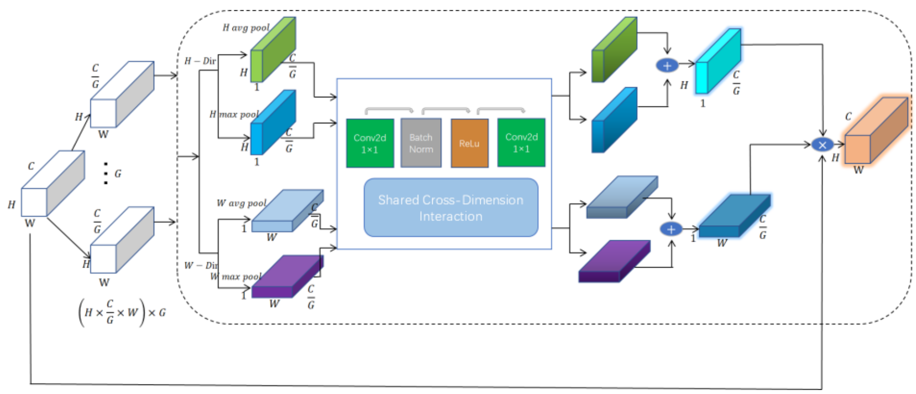

Fig. 2. GGCA Module

The GGCA module we have proposed is equipped with shared convolutional layers and a sophisticated attenuation mechanism, enabling it to intelligently generate attention maps along both the height and width dimensions. These attention maps serve as a precise weight-allocation mechanism, meticulously recalibrating the input feature maps of crucial forest fire characteristics.

In comparison to the SE module, the GGCA module captures multi-dimensional global information by conducting global average pooling and global max pooling separately along the height and width dimensions, thereby enhancing the comprehensiveness of feature extraction. Unlike CBAM, our GGCA module adopts a grouped processing strategy, grouping the input feature maps based on channels. This approach not only preserves the diversity and richness of feature representations but also reduces the computational overhead for each group.

3.3 EISDH

As illustrated in Figure 3(a), common detection heads employ two parallel branches to process class features and location features separately, achieving classification and localization through their respective 1×1 convolutional layers. Although this approach enhances the model's adaptability to complex scenarios and improves detection accuracy, it also leads to an increase in the number of parameters and computational load. To address this issue, we propose the Efficient Integrated Shared Detection Head (EISDH for short). We observe that the original model utilizes two 3×3 convolutional layers in each parallel branch. This design necessitates independent computation and weight storage for each branch, thereby increasing the model's parameter count and computational burden. To tackle this problem, as shown in Figure 3(b), we merge the 3×3 convolutional layers in the parallel branches into a single branch. The design of the shared convolutional layer significantly reduces the number of model parameters, simplifies the model architecture, and enhances the richness of feature information by integrating feature information from the parallel processing branches. This integration improves the effectiveness of smoke feature detection and ensures detection accuracy. By introducing residual connections, we mitigate information loss during transmission. Furthermore, to enhance the accuracy of smoke object localization and classification, we replace the traditional Batch Normalization (BN) with Group Normalization (GN), enabling more precise capture of smoke location features. Additionally, to accommodate multi-scale recognition of smoke features, we introduce a Scale layer for adaptive feature scaling, further improving detection accuracy and efficiency.

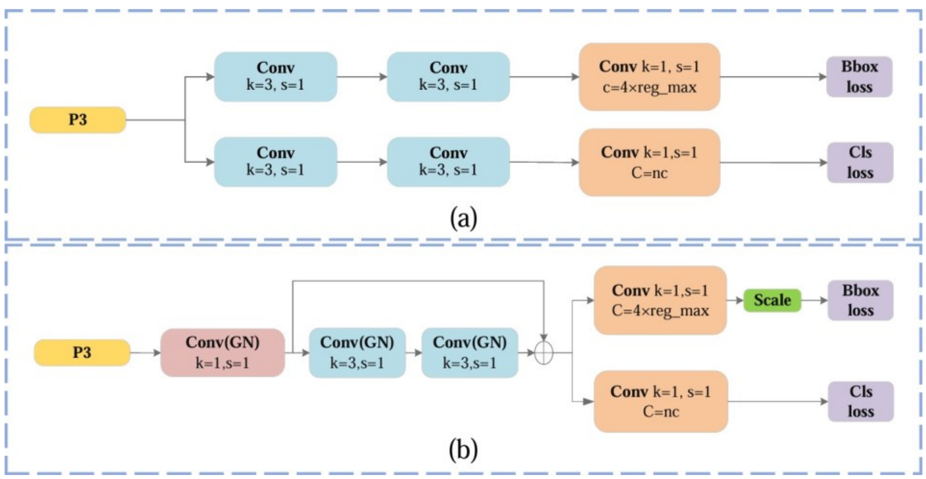

Fig. 3. (a) Decoupled Head, (b) EISDH Structure Diagram

EISDH enables our proposed smoke detection model to be not only more efficient but also more accurate and stable when handling smoke detection tasks, particularly in complex and ever-changing real-world scenarios, making the model more practical and reliable in real-world applications.

3.4 MPDIoU

In the object detection task for forest fires, the design of the loss function significantly impacts the accuracy of detection results. Its primary role is to optimize the positional discrepancy between predicted bounding boxes and ground-truth annotations, enabling the model to generate predictions that more closely ap- proximate the true object positions, thereby enhancing detection precision. The YOLOv10 model employs the Complete Intersection over Union (CIoU) loss function. However, due to its relatively ambiguous handling of aspect ratio variations, this function may lead to suboptimal optimization outcomes. Additionally, in practical detection scenarios, the significant variability among samples poses challenges for the model's adaptive adjustments to diverse detection targets, thus slowing down the convergence speed.

To address these limitations, this study replaces CIoU with the Minimum Point Distance-based Intersection over Union (MPDIoU) [15] loss function. By redesigning the loss metric to incorporate the minimum Euclidean distance be- tween the diagonal vertices (i.e., the top-left and bottom-right corners) of the predicted and ground-truth bounding boxes, MPDIoU effectively mitigates detection distortions caused by high sample variability and improves the model's robustness. This approach reduces the degrees of freedom in the loss function, simplifies the optimization process, and enhances adaptability to forest fire scenarios.

4 EXPERIMENT AND ANALYSIS

4.1 Introduction to the Dataset

The FFES dataset has been meticulously curated to include 6,546 images. During the construction process, we strictly adhered to the 8:1:1 ratio principle, scientifically partitioning the images into training, validation, and testing sets to ensure that the data can perform optimally during model training, validation, and testing phases. These images, sourced from a wide variety of online channels, have all undergone meticulous manual annotation by professionals be- fore being incorporated into the dataset, thereby guaranteeing the accuracy and consistency of the data annotations.

To construct a rich, diverse, and broadly representative dataset, we placed special emphasis on the diversity of scenes when selecting images. We specifically chose a series of smoke images from various scenarios, including numerous challenging samples. These samples encompass early forest fire smoke images captured under different weather conditions (such as sunny, cloudy, rainy, etc.), lighting environments (such as strong light, weak light, backlight, etc.), and from multiple angles (such as frontal, side, top-down views, etc.). Through these di- verse samples, we aim to more comprehensively simulate the complexities of the real world, thereby enhancing the model's generalization ability and robustness. Partial image examples can be seen in Figure 4.

Fig. 4. Schematic diagram of partial dataset

4.2 Experimental Details

The computational resource used for the experiment was a Tesla T4 GPU, equipped with an Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50GHz. The deep learning framework was Pytorch 3.10, with Python version 3.10 and CUDA version 12.2. The experimental parameters were set as follows: the input image size was 640×640, the batch size was 16, and the initial learning rate was set to 0.01. All other experiments were trained for 300 epochs.

4.3 Comparative Experiments

We conducted a comprehensive evaluation of the GEMN model on three bench- mark datasets, with a particular focus on lightweight metrics, including the number of parameters (Params), floating-point operations (GFLOPs), inference speed (FPS), and model size (Model Size). Meanwhile, we also paid attention to accuracy metrics such as precision, recall, and mean average precision at a 50% intersection over union (IoU) threshold (mAP50).

To verify the practicality and generalization ability of our improved algorithm, we employed two different public datasets for testing. The first dataset is Foggia1, which was used by EDMUNDO CASAS et al. [16]. This dataset comprises 8,974 images, with 3,731 images depicting fire scenes and the remaining 6,791 images portraying early-stage smoke scenes, making it particularly suit- able for wildfire and smoke detection. The second dataset is WildfireSmokeV12, a labeled dataset created by AI For Mankind3 based on publicly available HP- WREN4 camera images. Focused on smoke detection, it contains 744 images of pure smoke.

As shown in Table 1, on both public datasets, our model achieved improvements in both accuracy and mAP50, indicating that the GEMN model maintains good accuracy and demonstrates excellent generalization ability.

Table 1. Comparative experiments with public datasets

Dataset	Model	Precision	mAP50
Wildfire Smoke V1	YOLOv10n	0.920	0.945
	GEMN	**0.939**	**0.972**
Foggia	YOLOv10n	0.952	0.961
	GEMN	**0.967**	**0.974**

To further substantiate the performance advantages of the proposed method and meet the requirements for accuracy and speed in detecting critical components within real-world application scenarios, while also taking into account the computational constraints of hardware devices, we conducted comparative experiments under identical experimental conditions and parameter configurations. Specifically, we compared the GEMN deep neural network with current mainstream object detection algorithms, as well as several lightweight models improved upon YOLOv10n. The training duration for all models was uniformly set to 300 epochs.

As illustrated in Table 2, when compared with other network architectures, our proposed model exhibits superior performance, achieving the shortest inference time and the smallest number of parameters. This characteristic makes it particularly well-suited for hardware platforms with limited resources. Concurrently, the model attains the highest level of detection accuracy. Through these rigorous experiments, we have compellingly demonstrated both the efficiency and the applicability of our enhanced model.

Table 2. Comparison of results of different models

Model	Precision	mAP50	GFLOPs(G)	FPS	Inference time(ms)	Paras(M)	Model size(MB)
YOLOv5n [17]	0.939	0.978	16.0	97	3.3	7.0	8.2
YOLOv6n [18]	0.903	0.892	44.2	109	4.7	3.7	8.3
YOLOv7-Tiny [19]	0.890	0.933	**13.2**	103	2.2	6.0	9.3
YOLOv8n-ShuffleNetV2 [20]	0.895	0.928	28.8	131	1.4	6.7	6.0
YOLOv8n-MobileNetV3 [21]	0.891	0.923	24.7	143	1.6	6.2	**5.7**
YOLOv10n	0.931	0.966	18.5	157	1.3	4.4	1.0
YOLOv10n-ADG [22]	0.938	0.972	19.3	162	1.3	4.3	15.7
YOLOv10n-HPDD [23]	0.943	0.979	21.3	172	1.2	4.6	12.4
YOLOv10n-RTPD [24]	0.955	0.978	22.9	163	1.2	4.6	11.2
YOLOv10n-EfficientViT [25]	0.948	0.987	20.3	169	1.7	4.8	9.3
YOLOv10n-GhostNet [26]	0.963	0.989	19.2	167	1.2	5.6	13.5
YOLOv10n-ConvNeXtV2 [27]	0.961	0.984	18.1	177	1.0	4.6	8.2
GEMN(Ours)	**0.968**	**0.991**	16.7	**185**	**0.8**	**4.0**	6.9

To validate the reliability of our proposed method, we conducted extensive ablation experiments on the FFFS dataset.

As shown in Table 3, the detection accuracy of the unimproved YOLOv10 network is 93.1%, with a detection time of 1.3 milliseconds per image.

Table 3. Model Improvement ablation experiment

C2f-GA	EISDH	MPDIoU	Precision	Recall	mAP50	Inference time(ms)
×	×	×	0.931	0.951	0.966	1.3
✓	×	×	0.952	0.969	0.975	1.2
×	✓	×	0.939	0.958	0.977	**0.8**
×	×	✓	0.942	0.972	0.980	1.0
✓	✓	×	0.957	0.970	0.984	1.1
✓	×	✓	0.955	0.968	0.989	1.1
×	✓	✓	0.949	0.974	0.990	1.0
✓	✓	✓	**0.968**	**0.977**	**0.991**	0.9

After introducing our proposed C2f-GA module, the network's detection ac- curacy is further improved by 2.1%, and the detection time is reduced by 0.1 milliseconds. This enhancement stems from the optimized feature fusion in the model's intermediate layers, which improves multi-scale object recognition and is particularly beneficial for small object detection. The architecture demonstrates enhanced localization capabilities for key forest fire-related objects through re- fined feature aggregation. In the GGCA network, the deep integration of multidimensional global information with an advanced attention mechanism not only achieves accurate capture of multidimensional global information but also further strengthens feature representation, effectively reducing computational redundancy and model parameters while improving detection efficiency and preserving detailed feature representations.

After adopting EISDH, although the improvement in detection accuracy is not significant, the detection time is astonishingly reduced by 0.5 milliseconds. This is because EISDH replaces the original detection head in the YOLOv10 backbone network, reducing the model's complexity to some extent and thus significantly shortening the detection time. This makes the model more suitable for farmland pest and disease detection scenarios, still performing well on embedded systems with limited hardware resources.

Finally, we replaced CIoU with MPDIoU in the model, effectively mitigating bounding box distortions caused by large sample variability, improving the model's robustness, and reducing computational load and memory consumption. To balance the relationship between computational cost and detection performance, we further reduced the number of network model parameters.

Compared to the unimproved version, our proposed GEMN network achieves a cumulative improvement in detection accuracy of 3.7%, with enhanced detection performance for small objects in key areas. Additionally, the detection time is reduced by 0.4 milliseconds, further enhancing the network's detection rate. Moreover, our GEMN model excels in other metrics, exhibiting the shortest inference time and the fewest parameters. This characteristic makes it particularly suitable for hardware platforms with constrained resources. Simultaneously, the model attains the highest level of detection accuracy. Through these rigorous experiments, we have compellingly demonstrated both the efficiency and the applicability of our enhanced model.

5 CONCLUSIONS

This paper proposes a lightweight smoke detection network called GEMN, which enhances the algorithm's robustness and detection accuracy while reducing the model's computational complexity and the number of parameters, thereby improving the performance of smoke detection in complex scenarios. Firstly, to reduce the extraction of redundant features, we innovatively designed a GGCA attention mechanism. This mechanism significantly improves the comprehensiveness of the model's feature extraction by emphasizing the expression of important features. Secondly, to lower the model's computational complexity and parameter count, we introduced a lightweight detection head named EISDH. Addition- ally, we incorporated the MPDIoU function. This function not only enhances the model's robustness but also simplifies the process of extracting unnecessary features from forest fire targets, further reducing the number of parameters in the network model. Experimental results demonstrate that GEMN exhibits out- standing performance in smoke detection tasks within complex scenarios.

References

1. Ghali, R., Akhloufi, M.A.: Deep learning approaches for wildland fires remote sensing: Classification, detection, and segmentation. Remote Sensing **15**(7), 1821 (2023)
2. Alamgir, N., Nguyen, K., Chandran, V., Boles, W.: Combining multi-channel color space with local binary co-occurrence feature descriptors for accurate smoke detection from surveillance videos. Fire safety journal **102**, 1–10 (2018)
3. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence **24**(7), 971–987 (2002)
4. Liu, E.: Research on video smoke recognition based on dynamic image segmentation detection technology. In: 2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA). pp. 240–243. IEEE (2019)
5. Zeng, J., Lin, Z., Qi, C., Zhao, X., Wang, F.: An improved object detection method based on deep convolution neural network for smoke detection. In: 2018 international conference on machine learning and cybernetics (ICMLC). vol. 1, pp. 184–189. IEEE (2018)

6. Yang, H., Jang, H., Kim, T., Lee, B.: Non-temporal lightweight fire detection net- work for intelligent surveillance systems. IEEE Access **7**, 169257–169266 (2019)
7. Harjoseputro, Y., Yuda, I., Danukusumo, K.P., et al.: Mobilenets: Efficient convolutional neural network for identification of protected birds. IJASEIT (International Journal on Advanced Science, Engineering and Information Technology) **10**(6), 2290–2296 (2020)
8. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015)
9. Wang, G., Li, J., Zheng, Y., Long, Q., Gu, W.: Forest smoke detection based on deep learning and background modeling. In: 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS). pp. 112–116. IEEE (2020)
10. Li, Y., Shen, Z., Li, J., Xu, Z.: A deep learning method based on srn-yolo for forest fire detection. In: 2022 5th International Symposium on Autonomous Systems (ISAS). pp. 1–6. IEEE (2022)
11. Wang, S., Chen, T., Lv, X., Zhao, J., Zou, X., Zhao, X., Xiao, M., Wei, H.: Forest fire detection based on lightweight yolo. In: 2021 33rd Chinese Control and Decision Conference (CCDC). pp. 1560–1565. IEEE (2021)
12. Xue, Q., Lin, H., Wang, F.: Fcdm: an improved forest fire classification and detection model based on yolov5. Forests **13**(12), 2129 (2022)
13. Xu, H., Li, B., Zhong, F.: Light-yolov5: A lightweight algorithm for improved yolov5 in complex fire scenarios. Applied Sciences **12**(23), 12312 (2022)
14. Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., Ding, G.: Yolov10: Real- time end-to-end object detection. arXiv preprint arXiv:2405.14458 (2024)
15. Ma, S., Xu, Y.: Mpdiou: a loss for efficient and accurate bounding box regression. arXiv preprint arXiv:2307.07662 (2023)
16. Chen, J., Kao, S.h., He, H., Zhuo, W., Wen, S., Lee, C.H., Chan, S.H.G.: Run, don't walk: chasing higher flops for faster neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12021– 12031 (2023)
17. Fang, Y., Guo, X., Chen, K., Zhou, Z., Ye, Q.: Accurate and automated detection of surface knots on sawn timbers using yolo-v5 model. BioResources **16**(3), 5390 (2021)
18. Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W.e.a.: Yolov6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976 (2022)
19. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7464–7475 (2023)
20. Ma, N., Zhang, X., Zheng, H.T., Sun, J.: Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: Proceedings of the European conference on computer vision (ECCV). pp. 116–131 (2018)
21. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al.: Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1314–1324 (2019)
22. Zhou, K., Shi, J.L., Fu, J.Y., Zhang, S.X., Liao, T., Yang, C.Q., Wu, J.R., He, Y.C.:An improved yolov10 algorithm for automated damage detection of glass curtain- walls in high-rise buildings. Journal of Building Engineering p. 111812 (2025)
23. Aktouf, L., Shivanna, Y., Dhimish, M.: High-precision defect detection in solar cells using yolov10 deep learning model. In: Solar. vol. 4, pp. 639–659. MDPI (2024)

24. Li, Y., Leong, W., Zhang, H.: Yolov10-based real-time pedestrian detection for autonomous vehicles. In: 2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA). pp. 1–6. IEEE (2024)
25. Liu, X., Peng, H., Zheng, N., Yang, Y., Hu, H., Yuan, Y.: Efficientvit: Memory efficient vision transformer with cascaded group attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14420– 14430 (2023)
26. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C.: Ghostnet: More features from cheap operations. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1580–1589 (2020)
27. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S., Xie, S.: Convnext v2: Co-designing and scaling convnets with masked autoencoders. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16133–16142 (2023)