Open Journal of Applied Sciences, 2024, 14, 159-170
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

YOLOv8 for Fire and Smoke Recognition Algorithm Integrated with the Convolutional Block Attention Module

Zhangchi Liu[1], Risheng Zhang[2]*, Hao Zhong[1], Yingjie Sun[1]

[1]College of Railway Transportation, Hunan University of Technology, Zhuzhou, China
[2]Zhuzhou Boyan Intelligent Equipment Co. Ltd., Zhuzhou, China
Email: *liu_mu2000@163.com

How to cite this paper: Liu, Z.C., Zhang, R.S., Zhong, H. and Sun, Y.J. (2024) YOLOv8 for Fire and Smoke Recognition Algorithm Integrated with the Convolutional Block Attention Module. *Open Journal of Applied Sciences*, **14**, 159-170.
https://doi.org/10.4236/ojapps.2024.141012

Received: December 18, 2023
Accepted: January 28, 2024
Published: January 31, 2024

Abstract

The complexity of fire and smoke in terms of shape, texture, and color presents significant challenges for accurate fire and smoke detection. To address this, a YOLOv8-based detection algorithm integrated with the Convolutional Block Attention Module (CBAM) has been developed. This algorithm initially employs the latest YOLOv8 for object recognition. Subsequently, the integration of CBAM enhances its feature extraction capabilities. Finally, the WIoU function is used to optimize the network's bounding box loss, facilitating rapid convergence. Experimental validation using a smoke and fire dataset demonstrated that the proposed algorithm achieved a 2.3% increase in smoke and fire detection accuracy, surpassing other state-of-the-art methods.

Keywords

Object Recognition, CBAM, WioU, State-of-the-Art Methods

1. Introduction

Fire and smoke have become significant threats due to their high frequency and destructive nature. Their rapid spread, particularly in combustible-dense areas such as residential zones, airports, and forests, poses a challenge for swift control. Consequently, timely and accurate fire detection is crucial for preventing large-scale disasters. Traditionally, research has focused on contact-based fire detection sensors like smoke, temperature, and particle sensors, which are cost-effective and easy to deploy. However, these systems are suitable mainly for small areas and have considerable limitations in larger settings. Since they require direct activation by fire temperature or smoke, there is a potential delay in optimal fire extinguishing time. Compared to sensor-based methods, vision-based fire detection offers numerous advantages, including rapid response, extensive coverage, and environmental robustness, leading to its increasing popularity.

In recent years, research on fire and smoke detection has predominantly focused on video image detection algorithms, primarily divided into two categories: traditional classifier-based and deep learning-based smoke and fire detection. The former approach initially employs feature extraction methods such as SIFT [1] and HOG [2] to extract characteristics of fires and smoke, including brightness, color, texture, and edges. These features are then fed into classifiers for training, ultimately utilizing classifiers like SVM, Bayesian networks, and BP neural networks to determine the presence of fires and smoke in images, as discussed in [3]. However, this methodology predominantly relies on manually crafted algorithms for extracting low-level image features, followed by optimization of the results. Consequently, this leads to significant time consumption, resulting in poor performance and slower real-time detection of fire and smoke. Furthermore, issues like occlusion and interference often result in numerous false positives and errors in background detection. Therefore, these methods are ineffective for timely and efficient detection and alarm signaling in the early stages of tunnel fires.

Deep learning-based fire detection algorithms excel in extracting more abstract and advanced features of fires and smoke, demonstrating superior performance compared to traditional classifier-based methods. These algorithms are characterized by their high efficiency and accuracy. Frizzi S. [4] introduces a convolutional neural network capable of automatically recognizing fires in videos. This network utilizes convolutional layers to extract features, pooling layers to reduce feature map dimensions and simplify computational complexity, and fully connects layers to amalgamate all features before outputting to a classifier. Compared to manual feature extraction methods, these algorithms significantly improve accuracy and speed. However, their reliance on two-dimensional convolution overlooks the dynamic characteristics of fires and smoke. Moreover, due to dataset limitations, they are primarily effective in recognizing only red fires.

Cao Y. [5] and D. Nguyen M. [6] explore the application of Recurrent Neural Networks (RNNs) in fire detection tasks, utilizing their ability to extract relationships between features of the same object across different frames, thereby offering long-term memory of video information. Long Short-Term Memory networks (LSTMs), a variant of RNNs, address the issue of vanishing gradients present in traditional RNN models. When applied to fire detection, LSTMs are capable of simultaneously extracting spatial and temporal features of flames and smoke. This dual extraction results in high accuracy and recall rates while meeting real-time processing requirements. However, LSTMs present challenges due to their numerous fully connected layers, extensive time spans, deep network architecture, and the computational demand of numerous parameters, making them difficult to train. Panagiotis *et al.* [7] propose a fire detection method using an enhanced Faster R-CNN [8], which employs multi-dimensional texture analysis for feature extraction. This approach enables more accurate recognition of various types of flame images and offers adaptability to noise and lighting variations. Despite these advantages, the complexity of the algorithm is increased due to the extensive texture feature extraction, and the two-stage nature of Faster R-CNN, involving candidate region generation, leads to high precision and accurate localization but at the cost of a complex model structure and slower detection speed.

The YOLO [9] series represents a benchmark in single-stage detection algorithms. Cao *et al.* [10] introduced a fire and smoke detection model named SE_RFB_YOLO, which is based on the YOLOv3 [11] framework. This model incorporates a channel-based attention mechanism that enhances detection efficiency. Additionally, Cai W *et al.* [12] developed a smoke detection model named YOLO-SMOKE by embedding an efficient channel attention mechanism into the YOLOv3 model and modifying the loss function and this approach enhances the accuracy and robustness of the algorithm.

Numerous studies have already demonstrated the superiority of the YOLO series algorithms in the detection of smoke and flames. The YOLOv8 algorithm represents a further advancement by the original creators of YOLOv5, building upon its predecessors. To enhance the accuracy and robustness of smoke detection, this paper introduces a modified version of this algorithm, YOLOv8-CBAM, which incorporates the CBAM [13] (Convolutional Block Attention Module) into YOLOv8. Experiments conducted on a smoke and flame dataset and comparative analyses with YOLOv5, YOLOv6, and YOLOv8 have shown that YOLOv8-CBAM achieves a 2.3% increase in accuracy for smoke and flame detection, surpassing the performance of other methods.

The structure of this paper is organized as follows: Section 1 provides an introduction, setting the stage for the study. Section 2 elaborates on the fundamental principles of the YOLOv8-CBAM network framework. Section 3 presents comparative experiments with other smoke and flame detection algorithms, demonstrating the superiority of the YOLOv8-CBAM network. Finally, Section 4 offers a summary of the content and findings of this paper.

2. YOLOv8-CABM

As illustrated in **Figure 1**, the YOLOv8-CBAM architecture integrates three CBAM (Convolutional Block Attention Module) units into the base structure of YOLOv8.

2.1. YOLOv8n

The YOLOv8 algorithm is primarily composed of three parts: Backbone, Neck, and Head, as depicted in **Figure 2**. The Backbone primarily consists of multiple modules such as CBS, C2f, and SPPF, which are responsible for feature extraction from images. CBS represents a simple convolutional layer. The C2f module, drawing inspiration from the C3 module in YOLOv5 and the ELAN concept in YOLOv7 [14], is designed to ensure a richer gradient flow of information while maintaining a reduced number of parameters; its structure is also shown in Figure 2. The Neck part facilitates the integration of high-resolution and high-semantic information by merging high-level and low-level features. Finally, the Head, composed of multiple detection heads, is responsible for decoupling the refined feature information from the Neck, determining the position and category of the target object. The Backbone and Neck extract feature information but are incapable of performing localization tasks, which is the primary function of the Head.

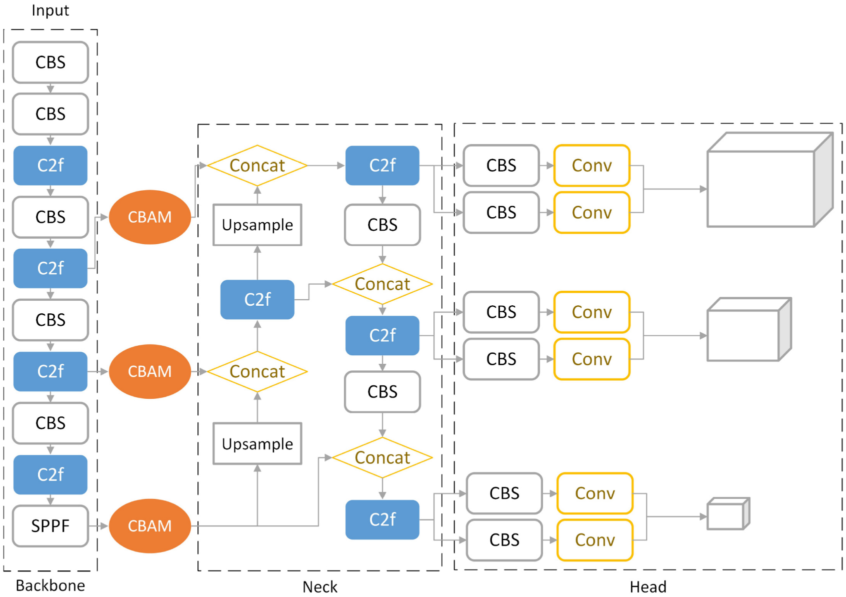

Figure 1. The structure of YOLOv8-CBAM.

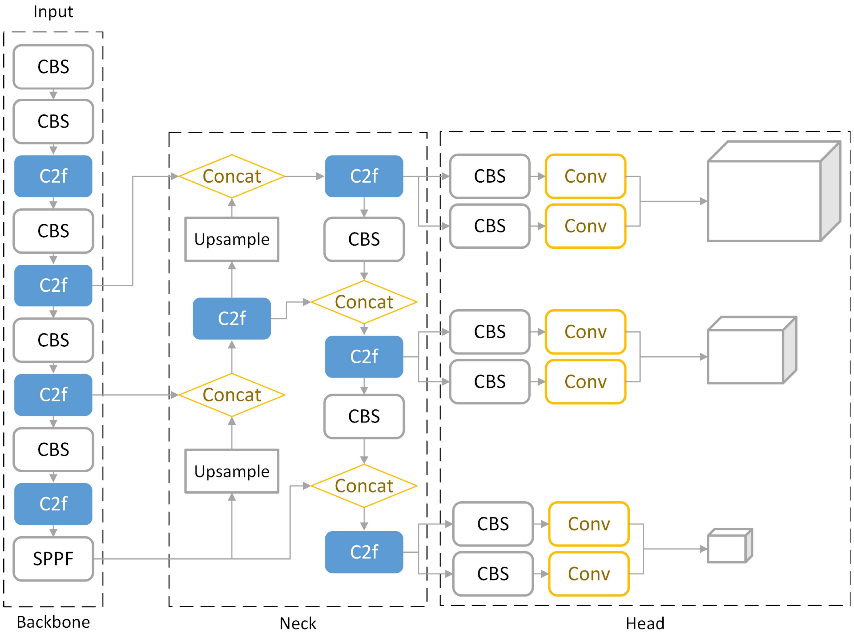

Figure 2. The structure of YOLOv8-CBAM.

2.2. CBAM

Given the often subtle and unstable movement characteristics of fires and smoke in certain scenarios, accurately detecting them poses a significant challenge for detection algorithms. To address this, the present study proposes the integration of the Convolutional Block Attention Module (CBAM) attention mechanism during the feature extraction phase of YOLOv8. CBAM combines channel and spatial attention mechanisms, effectively identifying key features in images while suppressing irrelevant noise. This dual attention mechanism notably enhances the accuracy and efficiency of detection, especially in complex and dynamic fire scenarios, making CBAM an essential tool in advanced image-based fire detection systems.

As depicted in Figure 3, CBAM consists of two modules: the Channel Attention Module (CAM), which implements channel attention mechanisms, and the Spatial Attention Module (SAM), which employs spatial attention mechanisms.

Figure 4 and Figure 5 respectively provide detailed illustrations of the basic structures of the Channel Attention Module (CAM) and the Spatial Attention Module (SAM).

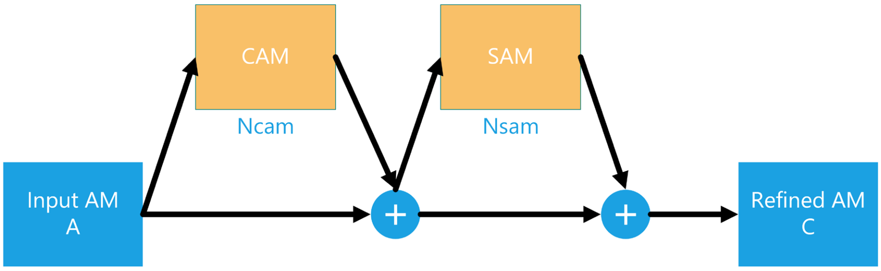

Figure 3. The structure of CBAM.

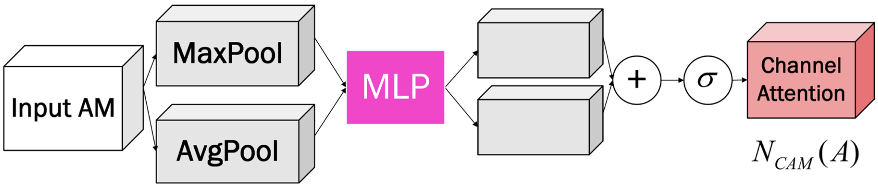

Figure 4. The structure of CAM.

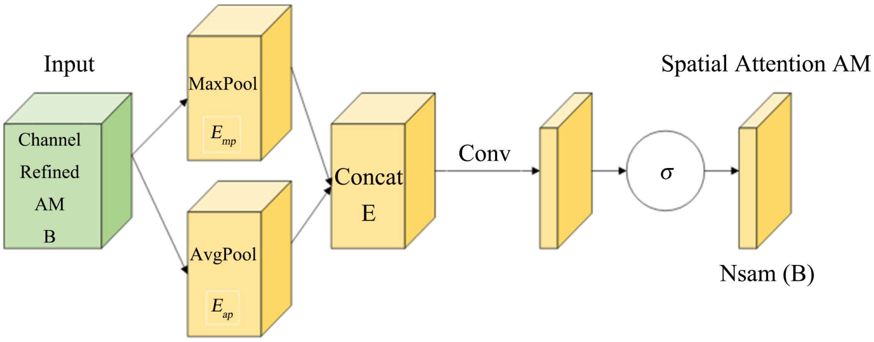

Figure 5. The structure of SAM.

Let the input feature map be denoted as. As illustrated in **Figure 5**, *A* first undergoes channel attention processing to obtain *B*, and then through spatial attention to yield the final activated feature map *C*. This process is mathematically represented in Equation (1):

$$\begin{cases} B = N_{CAM}(A) \otimes A \\ C = N_{SAM}(B) \otimes B \end{cases} \tag{1}$$

In this context, the symbol $\otimes$ represents element-wise multiplication. When the dimensions of the operands do not match, the spatial attention values are expanded along the channel dimension, while the channel attention values are expanded along the spatial dimensions.

2.3. Loss Function Optimization

The loss function of YOLOv8 comprises three components, as expressed in Equation (2):

$$L = L_{box} + L_{cls} + L_{DFL} \tag{2}$$

In this equation, L_{box}, L_{cls}, L_{DFL} represent the bounding box regression loss, classification loss, and Distribution Focal Loss (DFL), respectively. The bounding box regression loss is the Complete Intersection over Union (CIoU), with the full calculation detailed in Equation (3):

$$L_{CIoU} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v \tag{3}$$

In this context, α is a weighting function, v measures the similarity in aspect ratios, IoU is the Intersection over Union of the predicted and actual boxes, ρ denotes the Euclidean distance, *b* and b^{gt} are the center points of the actual and predicted boxes, respectively. *c* represents the diagonal length of the smallest enclosing box that contains both the predicted and actual boxes.

While CIoU effectively incorporates aspects such as distance, overlap area, center point deviation, and aspect ratio in bounding box regression, thus avoiding the issue present in DIoU where identical Intersection over Union (IoU) values cannot distinguish boxes with coinciding center points, it does not account for the directional mismatch between actual and predicted boxes. This paper opts to utilize WIoUv3 for bounding box regression loss. The computation formula for WIoUv1 loss is given in Equation (4):

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \tag{4}$$

The calculation formulas for R_{WIoU} and L_{IoU} are as Equation (5) and Equation (6):

$$R_{WIoU} = exp\left(\frac{\left(x - x_{gt}\right)^2 + \left(y - y_{gt}\right)^2}{\left(W_g^2 + H_g^2\right)^2}\right) \tag{5}$$

$$L_{IoU} = 1 - IoU \tag{6}$$

In these formulas, *x*, *y* and x_{gt}, y_{gt} respectively represent the center coordinates of the predicted and actual bounding boxes, while W_g and H_g denote the width and height of the actual bounding box. The calculation formulas for L_{WIoUv3} are as Equation (7), Equation (8) and Equation (9):

$$L_{WIoUv3} = rL_{WIoUv1} \tag{7}$$

$$r = \frac{\beta}{\delta \alpha^{\beta-\delta}} \tag{8}$$

$$\beta = \frac{L_{IoU}^{*}}{L_{IoU}} \in [0, +\infty) \tag{9}$$

3. Experimentation

3.1. Experimental Environment and Dataset

This study's experiments were conducted on a system running the Windows 11 operating system, powered by an Intel(R) Core i5-13490F CPU and an NVIDIA GeForce GTX 4070Ti GPU. The deep learning framework employed was PyTorch. After preparing the experimental dataset and setting up the experimental environment, iterative training was conducted using the proposed YOLOv8-CBAM model, along with other networks for comparative purposes. The dataset used in this study was a combination of the smoke public dataset mentioned in literature [15] and additional datasets collected through web scraping and publicly available online resources. This comprehensive dataset includes images of smoke and fires from various scenarios.

3.2. Experimental Evaluation Criteria

To accurately assess the model's effectiveness in detecting fires and smoke, this study employs precision, recall, mean Average Precision (mAP), and model forward inference time as key performance metrics.

- **Precision** evaluates the model's accuracy and is defined as the proportion of correct positive predictions out of all positive predictions made, as shown in Equation (10).
- **Recall** assesses the model's comprehensiveness by measuring the proportion of correct positive predictions out of all actual positive instances, as depicted in Equation (11).
- **mAP** is one of the most crucial performance evaluation metrics in the field of object detection, used to gauge the model's accuracy and comprehensiveness across multiple categories. The calculation process for mAP is outlined in Equation (12).

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

$$\text{mAp} = \frac{\sum_{n=1}^{N} \int_0^1 \text{Precision}_n \text{d}\left(\text{Recall}_n\right)}{N} \tag{12}$$

3.3. Experimental Results and Analysis

During training, the initial learning rate was set to 0.0001, with a batch size of 16 and the number of iterations fixed at 300. Both training and testing images were resized to a dimension of 640 × 640. **Figure 6** and **Figure 7** indicate that the model's training tended to stabilize after 100 iterations. Notably, during the final 10 iterations of training, the Mosaic augmentation was disabled, resulting in a significant downward trend in the curve. This demonstrates the effectiveness of the Mosaic augmentation in enhancing the model's performance.

3.3.1. CBAM Comparative Experiment

As previously mentioned, this paper integrates the CBAM polarized self-attention mechanism into the backbone network of YOLOv8n. To accurately evaluate the enhancement effect of CBAM on the existing algorithm, testing was extended beyond the original dataset to include images derived from real tunnel fire videos recorded in various complex scenarios, as depicted in **Figure 8**.

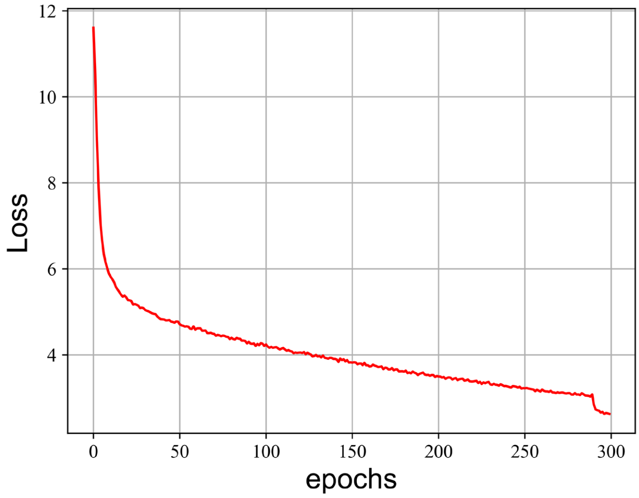

Figure 6. YOLOv8n loss curve.

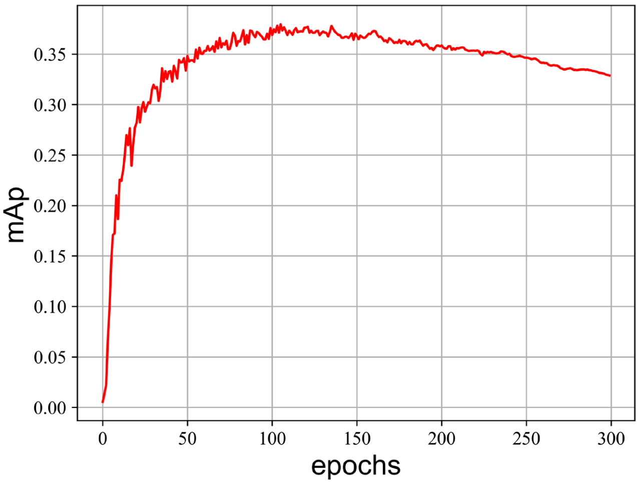

Figure 7. YOLOv8n mAP curve.

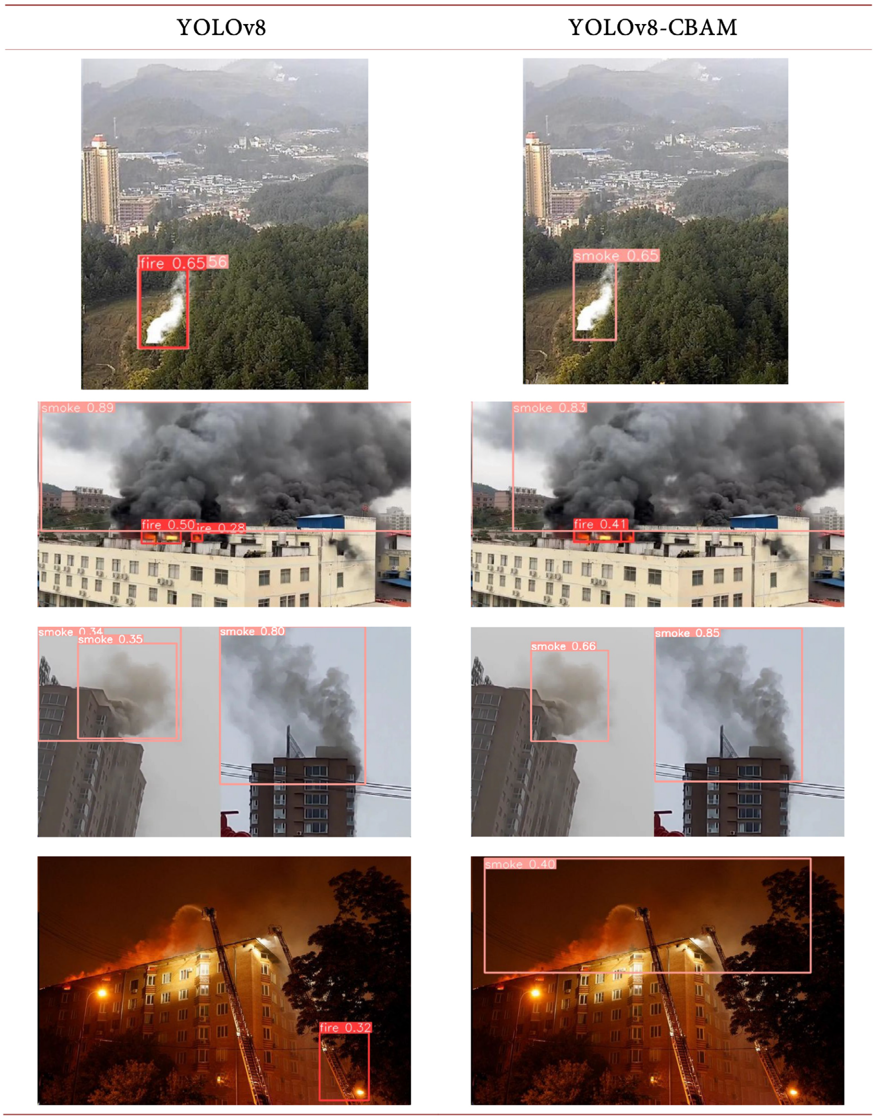

Figure 8. CBAM test results comparison.

3.3.2. Comparison Experiment with Other Models

To further evaluate the performance of the proposed method in smoke and fire detection, this study conducted a comparative analysis with widely used existing algorithms, including YOLOv5, YOLOv6, and the original YOLOv8. The results of this comparison are presented in Table 1. Compared to YOLOv5, YOLOv6, and YOLOv8, the proposed algorithm achieved a substantial improvement in accuracy for smoke and fire detection, with an increase of approximately 2.3 - 2.7 percentage points. Specifically, mean Average Precision at 50% IoU (mAP50) and mAP50-90 increased by 1.8 - 2.3 percentage points and 1.3 - 2 percentage points, respectively.

To provide a more visual demonstration of the model's performance, this paper selected four images for inference computation. As shown in Figure 9, each image represents a scenario with challenging or deceptive smoke and fire detection. The first image, depicting a sunset, was mistakenly identified as fire by

YOLOv5. The second image, correctly identifying smoke, was accurately recognized only by the model trained with YOLOv8-CBAM. In the third image, featuring multiple fires, other models either failed to detect them or produced overly large bounding boxes, lacking precision. The fourth image, representing a fiery sky, was incorrectly classified by all models except the proposed one. These instances clearly demonstrate the superiority of the algorithm proposed in this paper.

Table 1. Comparison of training results of different improved models.

	Precision	Recall	mAP50	mAP50-90
YOLOv5	0.78359	0.61919	0.66792	0.37307
YOLOv6	0.78172	0.632	0.66998	0.37956
YOLOv8	0.78542	0.60727	0.67293	0.37626
Our method	0.80917	0.63677	0. 69095	0.3933

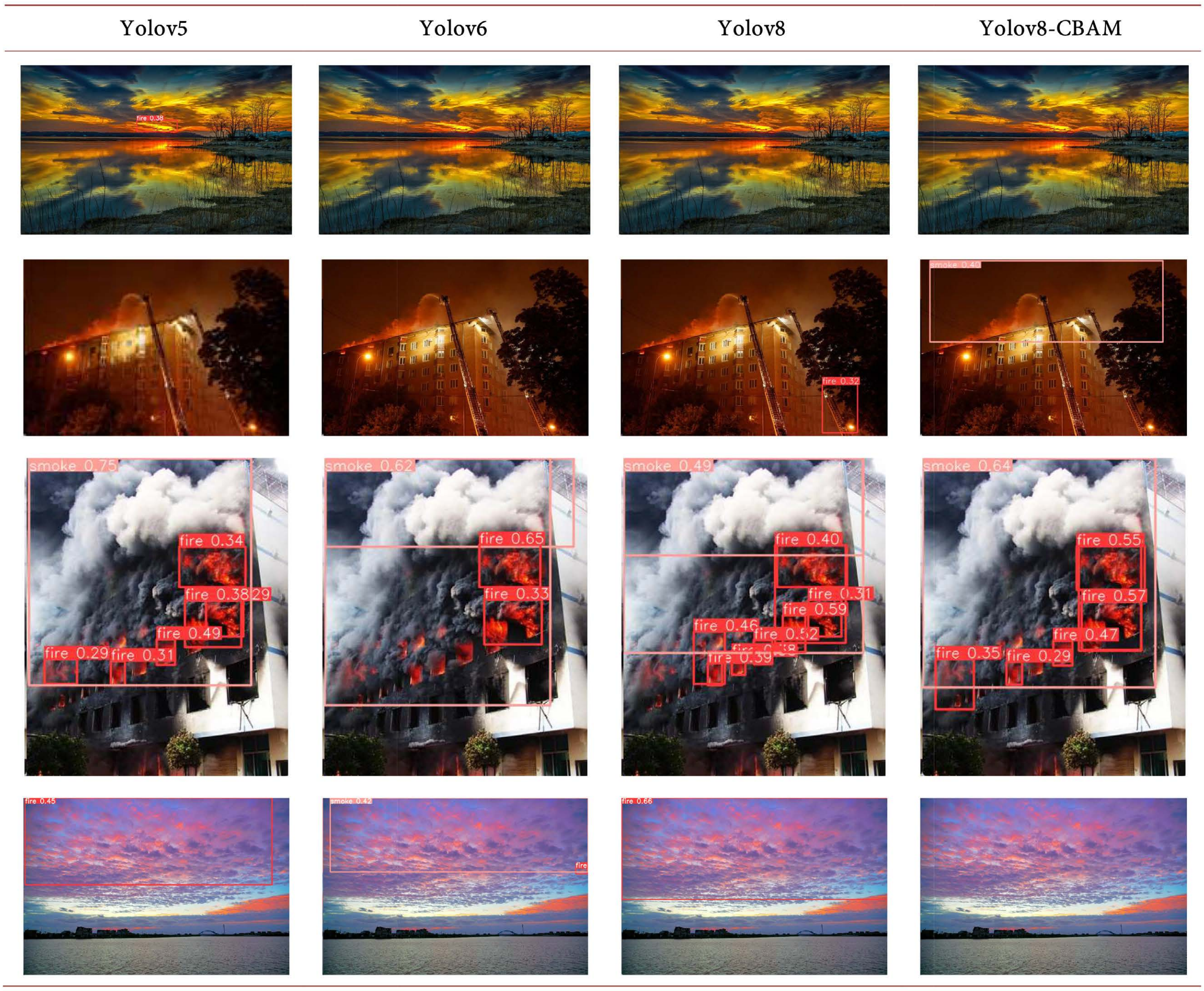

Figure 9. Comparison of detection results.

4. Conclusion

In this study, an enhanced smoke and fire detection algorithm based on the improved YOLOv8 framework and integrated with the Convolutional Block Attention Module (CBAM) demonstrated significant effectiveness in dealing with the complexities of shape, texture, and color in flames and smoke. The introduction of CBAM strengthened the algorithm's feature extraction capability, making the network more efficient in detecting two specific categories: smoke and fire. Additionally, the employment of the WIoU function optimized network loss and accelerated model convergence. Extensive training experiments conducted on a smoke and fire dataset indicated that the proposed algorithm substantially improved average precision compared to existing methods. However, the research also has limitations, such as the algorithm's adaptability in more complex smoke and fire scenarios not being fully validated. Future research will focus on exploring detection algorithms in more challenging smoke and fire environments to further validate and optimize the method proposed in this paper.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Lowe, D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[2] Kuang, H.L., Chan, L.L.H. and Yan, H. (2015) Multi-Class Fruit Detection Based on Multiple Color Channels. 2015 *International Conference on Wavelet Analysis and Pattern Recognition* (*ICWAPR*), Guangzhou, 12-15 July 2015, 9-15. https://doi.org/10.1109/ICWAPR.2015.7295917

[3] Dimitropoulos, K., Barmpoutis, P. and Grammalidis, N. (2015) Spatio-Temporal Flame Modeling and Dynamic Texture Analysis for Automatic Video-Based Fire Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, **25**, 339-351. https://doi.org/10.1109/TCSVT.2014.2339592

[4] Frizzi, S., Kaabi, R., Bouchouicha, M., Ginoux J.-M., Moreau, E. and Fnaiech, F. (2016) Convolutional Neural Network for Video Fire and Smoke Detection, *IECON* 2016 - 42*nd Annual Conference of the IEEE Industrial Electronics Society*, Florence, 23-26 October 2016, 877-882. https://doi.org/10.1109/IECON.2016.7793196

[5] Cao, Y., Yang, F., Tang, Q. and Lu, X. (2019) An Attention Enhanced Bidirectional LSTM for Early Forest Fire Smoke Recognition. *IEEE Access*, **7**, 154732-154742. https://doi.org/10.1109/ACCESS.2019.2946712

[6] Nguyen, M.D., Vu, H.N., Pham, D.C., Choi, B. and Ro, S. (2021) Multistage Real-Time Fire Detection Using Convolutional Neural Networks and Long Short-Term Memory Networks. *IEEE Access*, **9**, 146667-146679. https://doi.org/10.1109/ACCESS.2021.3122346

[7] Barmpoutis, P., Dimitropoulos, K., Kaza, K. and Grammalidis, N. (2019) Fire Detection from Images Using Faster R-CNN and Multidimensional Texture Analysis.

ICASSP 2019 - 2019 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*), Brighton, 12-17 May 2019, 8301-8305. https://doi.org/10.1109/ICASSP.2019.8682647

[8] Girshick, R. (2015) Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 7-13 December 2015, 1440-1448. https://doi.org/10.1109/ICCV.2015.169

[9] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, NV, USA, 27-30 June 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

[10] Cao, Y., Wang, G., Wen, H., Liu, X. and Yang, Z. (2022) Enhanced Receptive Field Smoke Detection Model Embedded with Attention Mechanism. 2022 *China Automation Congress* (*CAC*), Xiamen, 25-27 November 2022, 5122-5126. https://doi.org/10.1109/CAC57257.2022.10056099

[11] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 18-22 June 2018, 7794-7803.

[12] Cai, W., Wang, C., Huang, H. and Wang, T. (2020) A Real-Time Smoke Detection Model Based on YOLO-SMOKE Algorithm. 2020 *Cross Strait Radio Science & Wireless Technology Conference* (*CSRSWTC*), Fuzhou, 13-16 December 2020, 1-3. https://doi.org/10.1109/CSRSWTC50769.2020.9372453

[13] Woo, S., Park, J., Lee, J.-Y. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, Germany, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[14] Wang, C.-Y., Bochkovskiy, A. and Liao, H.-Y.M. (2023) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, 17-24 June 2023, 7464-7475. https://doi.org/10.1109/CVPR52729.2023.00721

[15] Yin, Z., Wan, B., Yuan, F., Xia, X. and Shi, J. (2017) A Deep Normalization and Convolutional Neural Network for Image Smoke Detection. *IEEE Access*, **5**, 18429-18438. https://doi.org/10.1109/ACCESS.2017.2747399