Article

A Flame Detection Algorithm Based on Improved YOLOv7

Guibao Yan [1], Jialin Guo [1], Dongyi Zhu [2], Shuming Zhang [1], Rui Xing [1], Zhangshu Xiao [1] and Qichao Wang [1,*]

[1]   College of Computer Science, Shaanxi Normal University, Xi'an 710119, China;
     yanguibao@snnu.edu.cn (G.Y.); guojl@snnu.edu.cn (J.G.); 1580533471@snnu.edu.cn (S.Z.);
     xingrui@snnu.edu.cn (R.X.); gregshaw@snnu.edu.cn (Z.X.)
[2]   School of International Studies, Shaanxi Normal University, Xi'an 710119, China; zhudy2018@snnu.edu.cn
*   Correspondence: wangqc@snnu.edu.cn

Abstract: Flame recognition is of great significance in fire prevention. However, current algorithms for flame detection have some problems, such as missing detection and false detection, and the detection accuracy cannot satisfy the requirements for fire prevention. In order to further the above problems, we propose a flame detection algorithm based on an improved YOLOv7 network. In our algorithm, we replace a convolution of the MP-1 module with a SimAM structure, which is a parameter-free attention mechanism. In this way, the missing detection problem can be improved. Furthermore, we use a ConvNeXt-based CNeB module to replace a convolution of the ELAN-W module for increasing detection accuracy and the false detection problem in complex environments. Finally, we evaluate the performance of our algorithm through a large number of test cases, and the data set used in our experiments was constructed by combining several publicly available data sets for various application scenarios. The experimental results indicate that compared with the original YOLOv7 algorithm, our proposed algorithm can achieve a 7% increase in the aspect of mAP_0.5 and a 4.1% increase in the aspect of F1 score.

Keywords: flame detection; deep learning; YOLO series; convolutional neural networks; target detection

1. Introduction

Fire disaster is one of the most common and widespread hazards to public safety, posing a threat to human life and property security. As a matter of fact, fire disasters can be easily extinguished in the early stages. Therefore, early detection of fire disasters plays an important role in fire prevention and may reduce economic losses and casualties. However, nowadays buildings are huge and complicated structures, which dramatically increases the difficulty of fire detection. Accordingly, a more advanced fire detection method is called for.

The development of technologies upon flame detection passes through stages such as sensor detection, traditional image processing methods, and deep learning methods. Given the importance of fire monitoring, extensive research has been carried out on flame detection methods. Traditional fire alarm systems generally detect changes in physical quantities such as smoke concentration and temperature in the environment [1]. When these physical quantities reach a certain value, an alarm is triggered, but there always exists a certain time lag. The current flame detection methods can be classified into two main categories: those based on manually defined flame features and those based on convolutional neural networks. Note that in the early stages of fire development, most flames are small and scattered, and their color is similar to natural light and sunlight. Therefore, the detection accuracy of the methods based on manually defined flame features is quite low, since it is difficult to detect small target flames. Researchers have tried to improve the detection accuracy of those methods by combining the flame's YUV color model, shape features, and motion features [2] and by combining the RGB color model with the ViBe background extraction algorithm [3]. Yet, the detection effect is still not satisfactory. Convolutional

neural networks have strong learning ability, fault tolerance, and fast speed; thus, they are commonly used in image recognition and classification. Currently, the convolutional neural networks (CNNs) used for object detection mainly include region-convolutional neural networks (R-CNN) [4] and YOLO series [5–9]. Compared with other convolutional neural networks, the YOLO series can better extract global information from images and can be trained end-to-end, which assures them as a more suitable option for flame detection. In recent years, the YOLO series has been developed over many generations [10–14]. However, current methods have some issues, such as low detection accuracy and high false-negative rate for small target flames, which cannot meet the requirements of flame prevention. YOLOv7 [15] is the latest version of the YOLO series, and in [16], it is proved that the performance of YOLOv7 is obviously better than YOLOv5 and YOLOv6 in the aspect of target detection. In earlier works, YOLOv7 has been applied to safety helmet detection, urban vehicle detection, road damage detection, tree species identification, and vehicle-related distance estimation [16–20]. In this work, we propose a new flame detection algorithm based on YOLOv7, which is, to our knowledge, the first time that YOLOv7 has been applied in this field. In our algorithm, we replace a convolution of the MP-1 module with the SimAM structure, which is a parameter-free attention mechanism [21]. In this way, the missing detection problem can be improved. Furthermore, we replace a convolution of the ELAN-W module with a ConvNeXt-based CNeB module [22] to improve detection accuracy and the false detection problem in complex environments. Additionally, in order to further the robustness of the algorithm, we construct a self-built data set by combining several publicly available data sets for various application scenarios, and thus it contains sufficient data volume and various complex detection environments. The experimental results demonstrate that our algorithm is distinctly stronger than the original YOLOv7 and YOLOv5 in all performance metrics.

This paper is structured as follows. In Section 2, we present the background and some related works, and in Section 3, we introduce the improvements of our proposed model based on YOLOv7. Then, we show our experiment design, including data set and performance metrics, in Section 4 and analyze the experiment results in Section 5. Finally, this paper is closed with a conclusion.

2. Background and Related Work

The goal of this section is to introduce the background of the field of target detection and some related works on parameter-free attention modules and ConvNeXt modules, which are applied in our proposed model.

2.1. Background on Target Detection

Target detection is a hot research topic in computer vision, and it can be divided into traditional target detection [23] and deep-learning-based target detection [24–28]. Traditional target detection methods mainly rely on feature extractors, which use sliding windows to extract image features and generate a large number of target candidate regions [29]. However, the above-mentioned methods are cumbersome and have some problems, such as serious window redundancy, slow detection speed, and low monitoring accuracy. Recently, convolutional neural networks have become widespread tools for image feature extraction and classification, which is the most popular architecture of deep learning [30–36]. The target detection technologies based on deep learning can adaptively learn high-level semantic information of images by using multi-structure network models along with their training algorithms.

In 2014, Girshick et al. successfully applied convolutional neural networks to target monitoring and proposed the R-CNN algorithm [4], which combined AlexNet [37] with selective search algorithms [38]. The detection accuracy of the R-CNN algorithm reached 58.5% on the PASCAL VOC2007 data set, which is significant progress compared to traditional target detection algorithms. The YOLO series is a deep-learning-based approach to object detection in real time. Meanwhile, it is often used in flame detection. In 2015, Redmon et al. proposed the YOLOv1 algorithm, which integrates classification, positioning, and detection functions in one network [10]. Following that, the YOLO series has been developed over many upgraded versions. There are many flame detection algorithms based on YOLOv3, YOLOv4, and YOLOv5 [5–9]. YOLOv7 [15] is the latest version of the YOLO series, and in [16], it is proved that the performance of YOLOv7 is obviously better than YOLOv5 and YOLOv6 in the aspect of target detection. Unfortunately, as far as we know, YOLOv7 has not yet been applied to flame detection. In this work, we propose an improved detection algorithm based on YOLOv7, which is, to our knowledge, the first time that YOLOv7 has been applied in this field.

2.2. Parameter-Free Attention Module

The attention mechanism is derived from the study of human vision, which means that the machine selectively focuses on a specific part of the visual area and ignores other irrelevant information. In recent years, the attention mechanism has been widely used in many areas, including image processing, speech recognition, and natural language processing [39–45]. In 2021, Yang et al. proposed the parameter-free attention module (SimAM) [21]. Compared to ECA single-dimensional attention [46] and CBAM two-dimensional attention [47], the SimAm attention module does not need to add parameters to derive three-dimensional attention weights, and it is simple and efficient. The structure comparison is shown in Figure 1.
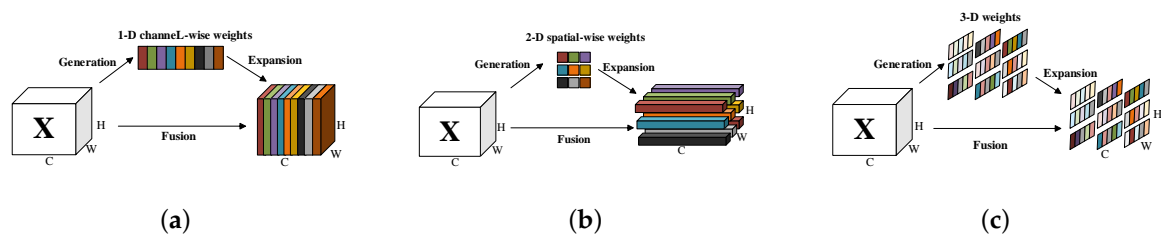

Figure 1. Comparisons of different attention steps: (**a**) Channel-wise attention; (**b**) Spatial-wise attention. (**c**) Full 3-D weights for attention.

In order to improve the performance of the attention mechanism, the SimAM algorithm evaluates the importance of each neuron. Rich information neurons inhibit surrounding neurons, known as spatial inhibition [48]. Informative neurons are found by measuring the linear separability between data elements, and an energy function is defined for each neuron in Equation (1).

$$e_t = (w_t, b_t, y, x_t) = (y_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2, \tag{1}$$

where

$$w_t = \frac{2(t - \mu_t)}{(t - \mu_t)^2 + 2\sigma_t^2 + 2\lambda},$$

$$b_t = -\frac{1}{2}(t + \mu_t)w_t,$$

$$\hat{t} = w_t t + b_t,$$

$$\hat{x}_i = w_t x_i + b_t,$$

t is the target neuron, and x_i are other neurons in a single channel of input features, respectively.

By minimizing Equation (1), the linear separability between neurons can be found [49]. For ease of understanding, y_t and y_0 are processed with binary labels, and a regularizer is added. Thus, we obtain the minimum energy from Equation (2)

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

where

$$\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i,$$

and

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

It is not difficult to see from Equation (2) that the smaller the energy of t is, the greater the linear separability between the target neuron and surrounding neurons is, and the higher the information richness is. Therefore, $\frac{1}{e_t^*}$ is generally used to indicate the importance of neurons.

2.3. ConvNeXt Model

ConvNeXt is a pure convolutional neural network proposed by Liu et al. [50]. It has a simple structure, and its accuracy and inference speed far exceed the Swin Transformer [51].

The structure is shown in Figure 2, where H, w, and *dim* represent the height, width, and number of layers of the feature map, respectively. The ConvNeXt network adjusts the stacking times in ResNet50 from $(3, 4, 6, 3)$ to $(3, 3, 9, 3)$, in order to obtain more complex features. First, the input feature map goes through a convolution layer with convolution kernel size 4 and step size 4 to downsample the input feature map. Then, it changes the number of channels to 96 and goes through 4 ConvNeXt blocks. Each ConvNeXt block adopts the inverted bottleneck design, which can effectively avoid the loss of information during the downsampling process, and then obtain a feature map with a size of $7 \times 7 \times 768$, and finally pass through the global average pool. The normalization and LN layers are used to reduce the model's dependence on the initialization parameters, such that the accuracy can be further improved, and finally, a feature map is output through a linear classifier.
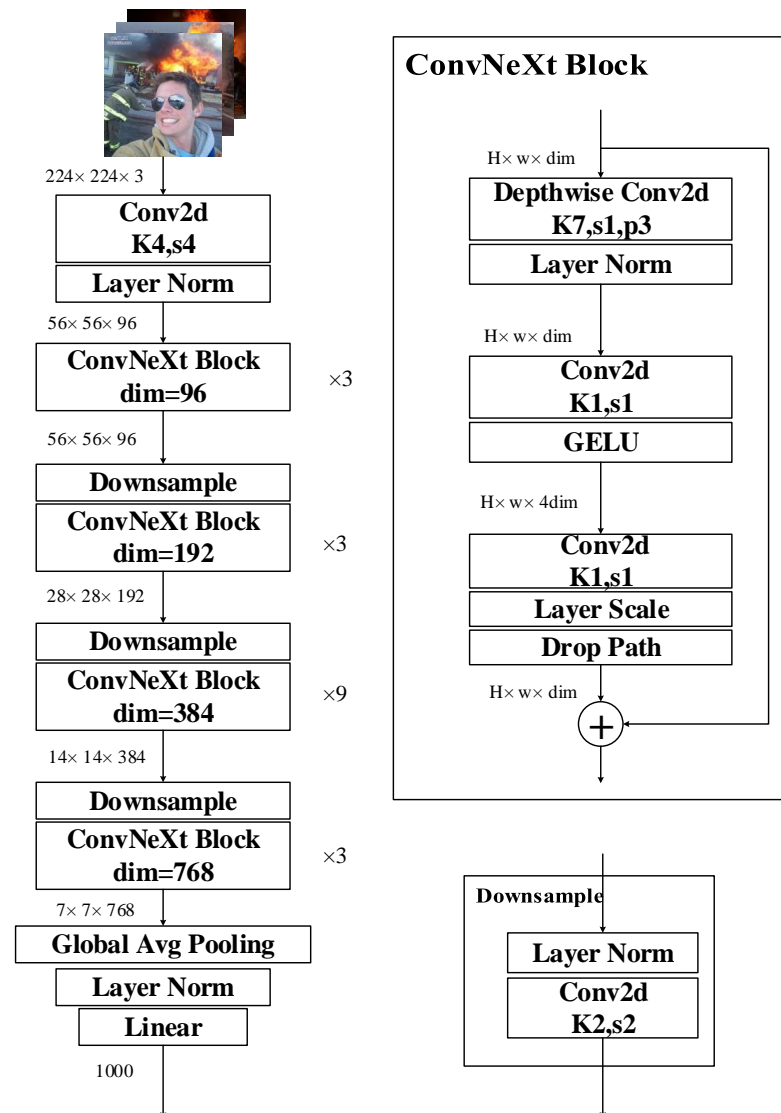
Figure 2. ConvNeXt network architecture.

3. Proposed Model

YOLOv7 is the most powerful target detection model of the YOLO series. Compared to earlier versions of the YOLO series, YOLOv7 is more efficient and more accurate, and it can reach a higher monitoring speed with the same computing resources.

For YOLOv7, the image is first resized to 640×640 and input to the backbone network. Then, three layers of feature maps of different sizes are output through the head layer network. Finally, prediction results are obtained through reparameterization and convolution. In general, YOLOv7 optimizes the model with the model structure reparameterization and dynamic label assignment. However, for non-rigid objects (e.g., flames), there are still deficiencies in the detection, such as missing detection, false detection, and low detection accuracy and efficiency. In order to further these problems, here, we propose an improved YOLOv7 network. In this work, we replace convolutions in MP-1 and ELAN-W modules of the YOLOv7 network with a three-dimensional attention mechanism SimAM and a pure ConvNeXt-based CNeB module, respectively. The newly added modules can be seen in the MP-1 and ELAN-W module structure diagrams in Figure 3 in detail.
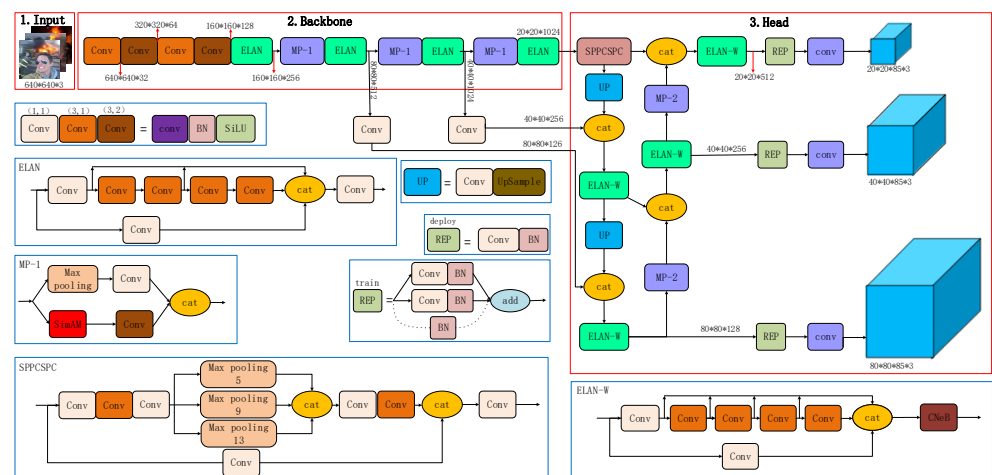
Figure 3. Improvement of YOLOv7 overall structure.

3.1. Improvement on YOLOv7 Backbone Structure

In the MP-1 module of the YOLOv7 network, the above branch goes through maximum pooling and 1×1 convolution, the image information is extracted through the local maximum, and during the below branch, the image information is extracted through two convolutions. When the information passes through the convolution with kernel 3×2, it will cause some fine-grained loss, thereby reducing the feature learning ability of the network, and thus small targets cannot be perceived. As shown in Figure 4, here, the 3×2 convolution in the below branch is replaced by the SimAM attention mechanism, which can suppress the interference of complex backgrounds on the target, enhance the extraction of target features, and significantly improve the problem of missing detection on small targets.
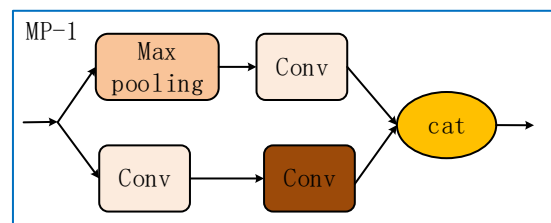
(**a**)
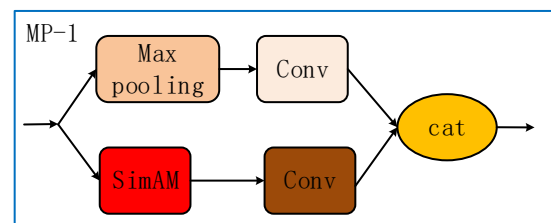
(**b**)

Figure 4. Comparisons of the original MP-1 module and the improved MP-1 module: (**a**) Original MP-1 module; (**b**) Improved MP-1 module.

3.2. Improvement on YOLOv7 Head Structure

In order to enable a deeper network to effectively learn and converge, the ELAN-W structure has been proposed in the version of YOLOv7. It is proved through experiments that by controlling the shortest and longest gradient path. The learning ability of the network can be improved without destroying the original gradient path.

As shown in Figure 5, here, we propose an improved ELAN-W module by replacing a 1×1 convolution with a CNeB module build based on ConvNeXt. The above branch first goes through a 1×1 convolution to change the number of channels. Then, the image features can be extracted through four 3×3 convolutions and finally passed to the CNeB module, which further improves the extraction of image features by the network model. There is a 1×1 convolution operation during the below branch, which is used to control the change in the number of channels. The improved ELAN-W module can further detection accuracy and the false detection problem.
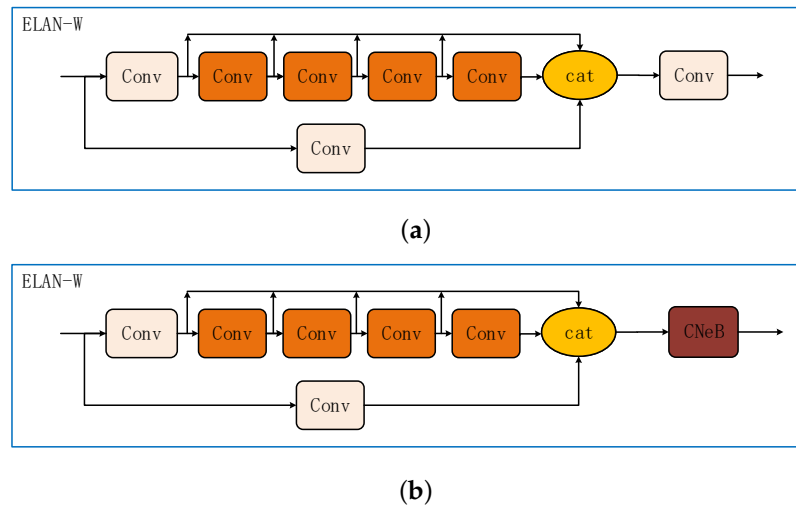

(**a**)

(**b**)

Figure 5. Comparisons of the original ELAN-W module and the improved ELAN-W module. (**a**) Original ELAN-W module; (**b**) Improved ELAN-W module.

4. Experiments Design

Our experiments were performed on the Ubuntu 16.04 operating system. The GPU and graphics card are NVIDIA TITAN Xp, and the display driver is version 470.74. Python version 3.7.13, CUDA version 12.1, and the PyTorch framework were used in programming for the experiments. In addition, the development environment is PyCharm.

4.1. Dataset

The data set plays a very crucial role in deep learning training, and it can make the model robust and generalizable. In order to satisfy the requirements of our experiments, we built a data set of flame under various scenes. The images used in the experiments were mainly from multiple public data sets of flames, such as Kaggle and ImageNet [52,53], and obtained by searching Google. Finally, we chose 8778 different effective images for our experiments, where the training set takes 70%, the testing set takes 20%, and validation set takes 10%, and some examples are shown in Figure 6.

Figure 6. Some example images of dataset.

4.2. Evaluation Metrics

The detection accuracy of this experiment is measured by the value of the *mean average precision* (*mAP*). This value represents the *average precision* (*AP*) of all categories, which is calculated from the *precision–recall curve* (*PR*) curve. The calculation formulas are as follows:

$$P = \frac{TP}{TP + FP}, \tag{3}$$

$$R = \frac{TP}{TP + FN}, \tag{4}$$

$$AP = \int_0^1 P(R)dR, \tag{5}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i, \tag{6}$$

where P indicates precision, R indicates recall, TP (true positives) is the number of correctly predicted positive samples, FP (false positives) is the number of wrongly predicted positive samples, FN (false negatives) is the number of wrongly predicted negative samples, and n represents the total number of target categories.

F1 score is another important performance metric that combines recall and precision together and takes both false positives and false negatives into account. The calculation formula of *F1* is as follows:

$$F1 = 2 \cdot \frac{P \cdot R}{P + R}. \tag{7}$$

5. Results and Analysis

In our experiments, we measured the performances of our improved model as well as the original YOLOv5 and YOLOv7, and the results are shown in Figure 7. In order to effectively compare these models, we used the same data set and the same training parameters and methods in the experiments for each model. In Figure 7, the so-called *YOLOv7-Improved-1* indicates the network only with the SimAM attention mechanism shown in Figure 4, the so-called *YOLOv7-Improved-2* indicates the network only with ConvNeXt shown in Figure 5, and the so-called *YOLOv7-Improved-3* indicates the network with both of SimAM attention mechanism and ConvNeXt. It can be easily seen from Figure 7 that there is no underfitting and overfitting during all experiments.

In order to avoid the randomness, the experiments for each model are repeated three times, and we choose the average value as the final result. An overview of our experimental results is given in Table 1. It can be easily seen that the overall performances of the original YOLOv7 has been significantly improved compared to YOLOv5. Further, the performances of YOLOv7-Improved-1 and YOLOv7-Improved-2 are already higher than the original YOLOv7 in each metric, and YOLOv7-Improved-3 is still much stronger than YOLOv7-Improved-1 and YOLOv7-Improved-2. The mAP_0.5 of the YOLOv7-Improved-3 has been increased by 7% compared to the original YOLOv7, the detection accuracy (precision) has been increased by 5.3%, and the F1 score has been increased by 4.1%.
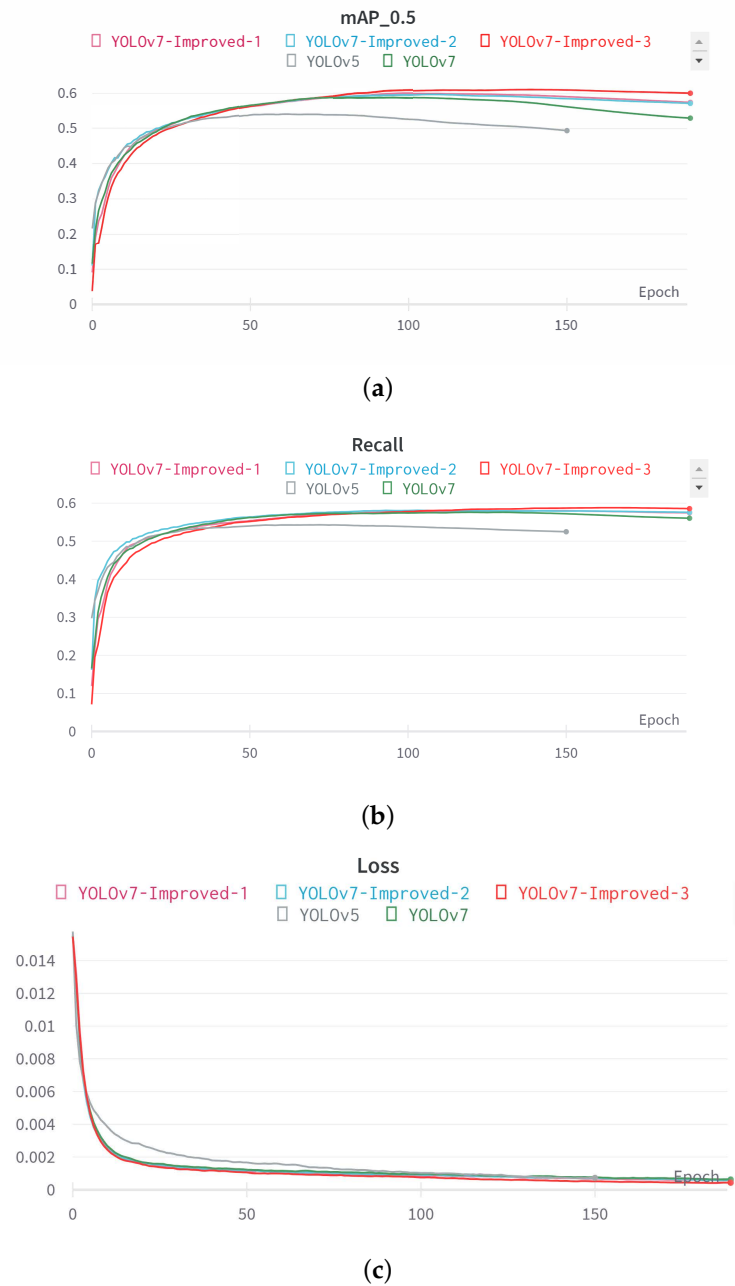
(**a**)

(**b**)

(**c**)

Figure 7. Comparisons of mAP, recall curves, and loss curves of various models: (**a**) mAP curves; (**b**) Recall curves; (**c**) Loss curves.

Table 1. Comparisons with the original YOLOv5 and YOLOv7.

Model	mAP_0.5	Precision	Recall	Parameters (M)	F1
YOLOv5	0.499	0.491	0.613	46.10	0.451
YOLOv7	0.544	0.543	0.621	36.40	0.560
YOLOv7-Improved-1	0.580	0.570	0.632	39.73	0.583
YOLOv7-Improved-2	0.585	0.577	0.630	40.52	0.576
YOLOv7-Improved-3	0.614	0.596	0.635	50.38	0.601

For comparative analysis of various models, we randomly selected three images from the test set and used the models of YOLOv5, YOLOv7, and YOLOv7-Improved-3 to detect the flame in these images. The detection effects are shown in Figure 8.
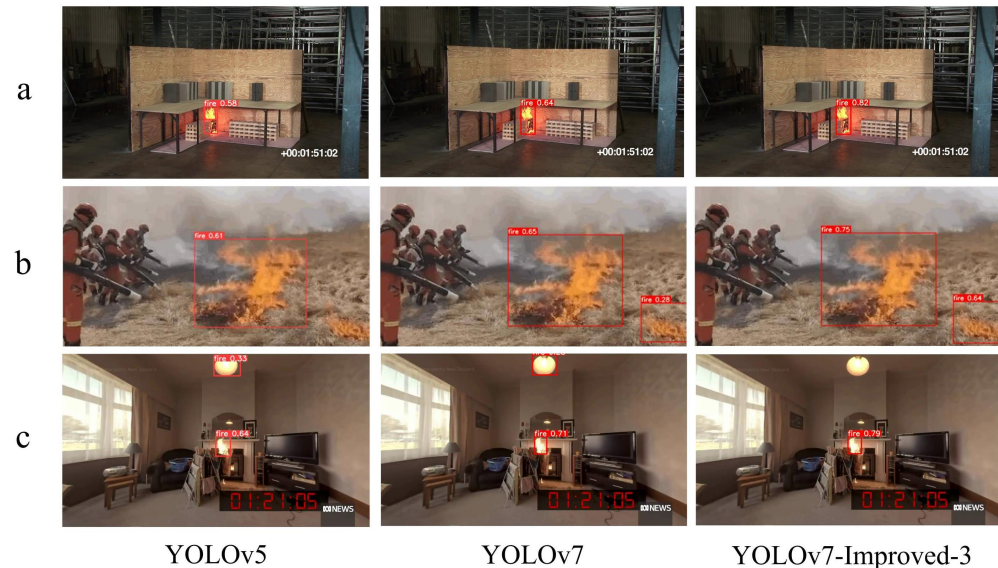

Figure 8. Comparison of test results.

From Figure 8, we can see that for Image a, the confidence of YOLOv5 is 58%, the confidence of YOLOv7 is 72%, and the confidence of YOLOv7-Improved-3 has reached 82%. Obviously, the confidence of YOLOv7-Improved-3 has been significantly improved. For Image b, YOLOv5 cannot detect the target in the lower right corner, while YOLOv7 and YOLOv7-Improved-3 detected it successfully, and the confidence of YOLOv7-Improved-3 is higher than YOLOv7. For Image c, both YOLOv5 and YOLOv7 have serious false detections due to the complex background and environment, while YOLOv7-Improved-3 detected the flame with a high degree of confidence.

6. Conclusions

In this work, we introduce a flame detection algorithm based on the improved YOLOv7 network. First, we added the SimAM non-parameter attention to the MP-1 module of YOLOv7 in order to solve the problem of missing detection of YOLOv7. Next, the ConvNeXt-based CNeB module replaces the ordinary convolution in the ELAN-W module of YOLOv7, which solves the false detection of YOLOv5 and YOLOv7 and also improves the performance of the model. Additionally, in order to enhance the robustness and generalization of our proposed algorithm, a sufficient amount of self-built data covering various application scenes is created based on different public data sets. The experimental results show that our improved YOLOv7 model has higher accuracy in flame detection without adding too many parameters, and the situations of missing detection and false detection have been distinctly improved. Although its detection speed is slower than the original YOLOv7, it can basically satisfy the requirement of real-time detection. In future work, the network structure will be further studied to improve the detection speed and the effect of real-time detection.

Author Contributions: Methodology, G.Y.; Software, S.Z. and R.X.; Formal analysis, G.Y.; Writing—original draft, G.Y., J.G. and D.Z.; Writing—review & editing, D.Z.; Supervision, Z.X. and Q.W.; Project administration, Z.X. and Q.W.; Funding acquisition, Q.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, L.; Wang, G. Design and Implementation of Automatic Fire Alarm System based on Wireless Sensor Networks. In Proceedings of the 2009 International Symposium on Information Processing, ISIP'09, Huangshan, China, 13–16 April 2009; pp. 410–413.
2. Foggia, P.; Saggese, A.; Vento, M. Real-Time Fire Detection for Video-Surveillance Applications Using a Combination of Experts Based on Color, Shape, and Motion. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 1545–1556. [CrossRef]
3. Zhang, Q.; Liu, X.; Huang, L. Video Image Fire Recognition Based on Color Space and Moving Object Detection. In Proceedings of the International Conference on Artificial Intelligence and Computer Engineering (ICAICE), Beijing, China, 23–25 October 2020.
4. Girshick, R.B.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, 23–28 June 2014; IEEE Computer Society: Washington, DC, USA, 2014; pp. 580–587. [CrossRef]
5. Wu, R.; Hua, C.; Ding, W.; Wang, Y.; Wang, Y. Flame and Smoke Detection Algorithm for UAV Based on Improved YOLOv4-Tiny. In Proceedings of the PRICAI 2021: Trends in Artificial Intelligence—18th Pacific Rim International Conference on Artificial Intelligence, PRICAI 2021, Hanoi, Vietnam, 8–12 November 2021; Pham, D.N., Theeramunkong, T., Governatori, G., Liu, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2021; Volume 13031, pp. 226–238. [CrossRef]
6. Zhang, H.; Wang, Z.; Chen, M.; Peng, Y.; Gao, Y.; Zhou, J. An Improved YOLOv3 Algorithm Combined with Attention Mechanism for Flame and Smoke Detection. In Proceedings of the Artificial Intelligence and Security—7th International Conference, ICAIS 2021, Dublin, Ireland, 19–23 July 2021; Sun, X., Zhang, X., Xia, Z., Bertino, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2021; Volume 12736, pp. 226–238. [CrossRef]
7. Wang, Y.; Hua, C.; Ding, W.; Wu, R. Real-time detection of flame and smoke using an improved YOLOv4 network. *Signal Image Video Process* **2022**, *16*, 1109–1116. [CrossRef]
8. Yang, T.; Xu, S.; Li, W.; Wang, H.; Shen, G.; Wang, Q. A Smoke and Flame Detection Method Using an Improved YOLOv5 Algorithm. In Proceedings of the IEEE International Conference on Real-Time Computing and Robotics, RCAR 2022, Guiyang, China, 17–22 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 366–371. [CrossRef]
9. Ma, J.; Zhang, Z.; Xiao, W.; Zhang, X.; Xiao, S. Flame and Smoke Detection Algorithm Based on ODConvBS-YOLOv5s. *IEEE Access* **2023**, *11*, 34005–34014. [CrossRef]
10. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; IEEE Computer Society: Washington, DC, USA, 2017; pp. 6517–6525.
11. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
12. Bochkovskiy, A.; Wang, C.; Liao, H.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
13. Kuznetsova, A.; Maleva, T.; Soloviev, V. Detecting Apples in Orchards Using YOLOv3 and YOLOv5 in General and Close-Up Images. In Proceedings of the Advances in Neural Networks—ISNN 2020—17th International Symposium on Neural Networks, ISNN 2020, Cairo, Egypt, 4–6 December 2020; Han, M., Qin, S., Zhang, N., Eds.; Springer: Berlin/Heidleberg, Germany, 2020; pp. 233–243.
14. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976.
15. Wang, C.; Bochkovskiy, A.; Liao, H.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:abs/2207.02696.
16. Chen, K.; Yan, G.; Zhang, M.; Xiao, Z.; Wang, Q. Safety Helmet Detection Based on YOLOv7. In Proceedings of the The 6th International Conference on Computer Science and Application Engineering, CSAE 2022, Virtual Event, China, 21–23 October 2022; Emrouznejad, A., Ed.; ACM: New York, NY, USA, 2022; pp. 31:1–31:6. [CrossRef]
17. Zhang, Y.; Sun, Y.; Wang, Z.; Jiang, Y. YOLOv7-RAR for Urban Vehicle Detection. *Sensors* **2023**, *23*, 1801. [CrossRef]
18. Liu, X.; Yan, W.Q. Vehicle-Related Distance Estimation Using Customized YOLOv7. In Proceedings of the Image and Vision Computing—37th International Conference, IVCNZ 2022, Auckland, New Zealand, 24–25 November 2022; Yan, W.Q., Nguyen, M., Stommel, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2022; Volume 13836, pp. 91–103. [CrossRef]
19. Hu, B.; Zhu, M.; Chen, L.; Huang, L.; Chen, P.; He, M. Tree species identification method based on improved YOLOv7. In Proceedings of the 8th IEEE International Conference on Cloud Computing and Intelligent Systems, CCIS 2022, Chengdu, China, 26–28 November 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 622–627. [CrossRef]
20. Pham, V.; Nguyen, D.; Donan, C. Road Damage Detection and Classification with YOLOv7. In Proceedings of the IEEE International Conference on Big Data, Big Data 2022, Osaka, Japan, 17–20 December 2022; Tsumoto, S., Ohsawa, Y., Chen, L., den Poel, D.V., Hu, X., Motomura, Y., Takagi, T., Wu, L., Xie, Y., Abe, A., et al., Eds.; IEEE: Piscataway, NJ, USA, 2022; pp. 6416–6423. [CrossRef]

21. Yang, L.; Zhang, R.; Li, L.; Xie, X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In Proceedings of the Proceedings of the 38th International Conference on Machine Learning, ICML 2021, Virtual Event, 18–24 July 2021; Meila, M., Zhang, T., Eds.; Proceedings of Machine Learning Research, PMLR: London, UK, 2021; Volume 139, pp. 11863–11874.

22. Zheng, Y.; Zhang, Y.; Qian, L.; Zhang, X.; Diao, S.; Liu, X.; Cao, J.; Huang, H. A lightweight ship target detection model based on improved YOLOv5s algorithm. *PLoS ONE* **2023**, *18*, e0283932. [CrossRef]

23. Felzenszwalb, P.F.; Girshick, R.B.; McAllester, D.A.; Ramanan, D. Object Detection with Discriminatively Trained Part-Based Models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1627–1645. [CrossRef]

24. Liu, D.; Teng, W. Deep learning-based image target detection and recognition of fractal feature fusion for BIOmetric authentication and monitoring. *Netw. Model. Anal. Health Inform. Bioinform.* **2022**, *11*, 17. [CrossRef]

25. Kumar, M.B.; Kumar, P.R. Moving Target Detection Strategy Using the Deep Learning Framework and Radar Signatures. *Int. J. Swarm Intell. Res.* **2022**, *13*, 1–21. [CrossRef]

26. Israsena, P.; Pan-Ngum, S. A CNN-Based Deep Learning Approach for SSVEP Detection Targeting Binaural Ear-EEG. *Frontiers Comput. Neurosci.* **2022**, *16*, 868642. [CrossRef]

27. Zheng, H.; Liu, J.; Ren, X. Dim Target Detection Method Based on Deep Learning in Complex Traffic Environment. *J. Grid Comput.* **2022**, *20*, 8. [CrossRef]

28. Jia, P.; Zheng, Y.; Wang, M.; Yang, Z. A deep learning based astronomical target detection framework for multi-colour photometry sky survey projects. *Astron. Comput.* **2023**, *42*, 100687. [CrossRef]

29. Viola, P.A.; Jones, M.J. Rapid Object Detection using a Boosted Cascade of Simple Features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), with CD-ROM, Kauai, HI, USA, 8–14 December 2001; IEEE Computer Society: Washington, DC, USA, 2001; pp. 511–518. [CrossRef]

30. Sisco, Y.J.; Carmona, R. Face recognition using deep learning feature injection: An accurate hybrid network combining neural networks based on feature extraction with convolutional neural network. In Proceedings of the XLVIII Latin American Computer Conference, CLEI 2022, Armenia, CO, USA, 17–21 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–11. [CrossRef]

31. Kashir, B.; Ragone, M.; Ramasubramanian, A.; Yurkiv, V.; Mashayek, F. Application of fully convolutional neural networks for feature extraction in fluid flow. *J. Vis.* **2021**, *24*, 771–785. [CrossRef]

32. Ezhilarasan, A.; Selvaraj, A.; Jebarani, W.S.L. MixNet: A Robust Mixture of Convolutional Neural Networks as Feature Extractors to Detect Stego Images Created by Content-Adaptive Steganography. *Neural Process. Lett.* **2022**, *54*, 853–870. [CrossRef]

33. Uzkent, B.; Yeh, C.; Ermon, S. Efficient Object Detection in Large Images Using Deep Reinforcement Learning. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, WACV 2020, Snowmass Village, CO, USA, 1–5 March 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1813–1822. [CrossRef]

34. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, 13–19 June 2020; Computer Vision Foundation/IEEE: Piscataway, NJ, USA, 2020; pp. 10778–10787. [CrossRef]

35. Cui, Y.; Yang, L.; Liu, D. Dynamic Proposals for Efficient Object Detection. *arXiv* **2022**, arXiv:2207.05252.

36. Cai, Y.; Luan, T.; Gao, H.; Wang, H.; Chen, L.; Li, Y.; Sotelo, M.Á.; Li, Z. YOLOv4-5D: An Effective and Efficient Object Detector for Autonomous Driving. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–13. [CrossRef]

37. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

38. Uijlings, J.R.R.; van de Sande, K.E.A.; Gevers, T.; Smeulders, A.W.M. Selective Search for Object Recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [CrossRef]

39. Han, J.; Huang, C.; Sun, S.; Liu, Z.; Liu, J. bjXnet: An improved bug localization model based on code property graph and attention mechanism. *Autom. Softw. Eng.* **2023**, *30*, 12. [CrossRef]

40. Duan, G.; Dong, Y.; Miao, J.; Huang, T. Position-Aware Attention Mechanism-Based Bi-graph for Dialogue Relation Extraction. *Cogn. Comput.* **2023**, *15*, 359–372. [CrossRef]

41. Ma, H.; Yang, K.; Pun, M. Cellular traffic prediction via deep state space models with attention mechanism. *Comput. Commun.* **2023**, *197*, 276–283. [CrossRef]

42. Chen, Z.; Li, J.; Liu, H.; Wang, X.; Wang, H.; Zheng, Q. Learning multi-scale features for speech emotion recognition with connection attention mechanism. *Expert Syst. Appl.* **2023**, *214*, 118943. [CrossRef]

43. Wang, W.; Li, Q.; Xie, J.; Hu, N.; Wang, Z.; Zhang, N. Research on emotional semantic retrieval of attention mechanism oriented to audio-visual synesthesia. *Neurocomputing* **2023**, *519*, 194–204. [CrossRef]

44. Ding, P.; Qian, H.; Zhou, Y.; Chu, S. Object detection method based on lightweight YOLOv4 and attention mechanism in security scenes. *J. Real Time Image Process.* **2023**, *20*, 34. [CrossRef]

45. Liu, W.; Lu, H.; Wang, Y.; Li, Y.; Qu, Z.; Li, Y. MMRAN: A novel model for finger vein recognition based on a residual attention mechanism. *Appl. Intell.* **2023**, *53*, 3273–3290. [CrossRef]

46. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, 13–19 June 2020; Computer Vision Foundation/IEEE: Piscataway, NJ, USA, 2020; pp. 11531–11539. [CrossRef]

47. Woo, S.; Park, J.; Lee, J.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the Computer Vision—ECCV 2018—15th European Conference, Munich, Germany, 8–14 September 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11211, pp. 3–19. [CrossRef]
48. Webb, B.S.; Dhruv, N.T.; Solomon, S.G.; Tailby, C.; Lennie, P. Early and late mechanisms of surround suppression in striate cortex of macaque. *J. Neurosci. Off. J. Soc. Neurosci.* **2006**, *25*, 11666–11675. [CrossRef]
49. Luo, F.; Zou, Z.; Liu, J.; Lin, Z. Dimensionality Reduction and Classification of Hyperspectral Image via Multistructure Unified Discriminative Embedding. *IEEE Trans. Geosci. Remote. Sens.* **2022**, *60*, 1–16. [CrossRef]
50. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Orleans, LA, USA, 18–24 June 2022; pp. 11966–11976. [CrossRef]
51. Hariharan, B.; Malik, J.; Ramanan, D. Discriminative Decorrelation for Clustering and Classification. In Proceedings of the Computer Vision—ECCV 2012, Florence, Italy, 7–13 October 2012; Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 459–472.
52. Kaggle. Available online: https://www.kaggle.com (accessed on 11 March 2023).
53. Image Net. Available online: https://image-net.org (accessed on 11 March 2023).