

## Article

# DCNFYOLO: Dual-Convolution Network and Feature Fusion for High-Precision Smoke Detection

Xin Chen <sup>1,2</sup> , Xuzhao Liu <sup>1,\*</sup> , Bing Liu <sup>3</sup> and Yaolin Zhu <sup>1</sup>

<sup>1</sup> School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China; chenxin@xpu.edu.cn (X.C.); fz\_zyl@126.com (Y.Z.)

<sup>2</sup> School of Automation, Northwestern Polytechnical University, Xi'an 710072, China

<sup>3</sup> Shaanxi Architectural Design Research Institute Co., Ltd., Xi'an 710018, China; 13571910851@163.com

\* Correspondence: 230421149@stu.xpu.edu.cn

**Abstract:** Fast, real-time, and accurate detection of smoke characteristics in the early stage of a fire is crucial for reducing fire losses. Existing smoke detection methods mainly rely on traditional algorithms and smoke sensors, and these approaches have limitations in false detection rates, accuracy, and real-time performance. Therefore, a novel DCNFYOLO network for smoke detection is proposed in this paper. Firstly, Switchable Atrous Convolution (SAConv) is introduced in the YOLOv5 backbone network to enhance the fusion extraction of smoke features by the Convolutional Neural Network (CNN). Secondly, both Distribution Shifts Convolution (DSConv) operator and Efficient Channel Attention (ECA) mechanisms are considered in the neck part to reduce the computational load of the model, and better capture the relationship between channels to improve the detection performance. Finally, to make low-quality examples less harmful to the gradients, the Wise-IoU (WIoU) loss function in the prediction part is used for reducing the competitiveness of high-quality anchor frames during model training, allowing the model to converge more quickly and stably. Experimental results show that the DCNFYOLO network can achieve a remarkable detection accuracy of 96.6%, which has a substantial improvement of 7.7% compared with the original YOLOv5 network performance, thereby validating the effectiveness of the proposed network.

**Keywords:** smoke detection; computer vision; SAConv; DSConv



**Citation:** Chen, X.; Liu, X.; Liu, B.; Zhu, Y. DCNFYOLO:

Dual-Convolution Network and Feature Fusion for High-Precision Smoke Detection. *Electronics* **2024**, *13*, 3864. <https://doi.org/10.3390/electronics13193864>

Academic Editor: Maciej Ławryńczuk

Received: 9 September 2024

Revised: 25 September 2024

Accepted: 27 September 2024

Published: 29 September 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Fire constantly endangers personal safety and social property [1], with more than 10,000 fires occurring worldwide daily, resulting in hundreds of deaths and direct property losses of over USD 1 billion. In order to reduce the frequency of fire, this paper proposes a deep learning-based smoke detection algorithm. At present, there are two primary methods for fire prevention: flame detection and smoke detection. Since smoke is often the initial indicator of a fire, identifying it promptly and accurately can effectively reduce the probability of fires. By detecting smoke and sounding an alarm in a timely manner, a quick response can be achieved to prevent a fire from spreading, thus avoiding casualties and property damage. Smoke detection is widely used in a variety of environments such as homes, office buildings, factories, and public places, and are a core component of fire safety.

In general, smoke detection mainly relies on traditional algorithms and sensors to collect smoke data. For example, widely used sensors for smoke detection are MH-410D, MH-711A, and MH-Z19B. Smoke sensors are crafted with user-friendly designs, ensuring effortless installation and operation without the need for complex setup procedures to initiate detection. However, the short-range communications, low accuracy, and susceptibility to external environments of sensors may lead to unstable detection results. Moreover, smoke sensors typically rely on smoke particles diffusing to the detector in order to trigger an alarm. As a result, they often respond slowly in large spaces or well-ventilated environments. This delay can lead to missing the optimal time for fire intervention, especially

in situations where the fire spreads rapidly, potentially resulting in serious consequences. Conventional smoke detection algorithms typically comprise machine learning techniques like Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes, etc. Compared with sensor detection methods, the detection range of machine learning-based smoke detection algorithms has been outperforming the existing schemes in terms of reliability, accuracy rate, and miss detection rate. But machine learning models are sensitive to environmental changes, such as the effects of lighting and background noise. When the environment changes dramatically, the model performance may decline. Additionally, the accuracy of the model depends on the training dataset, and the reliability of the model can be affected if the training data are insufficient or biased.

Given the swift advancement of computer vision and image processing, deep learning methods can dynamically acquire and discern smoke features from fires [2]. Compared with traditional target detection algorithms, deep learning-based object detection algorithms have strong robustness in complex scenes [3]. Deep learning models are able to learn complex features to improve the accuracy of smoke detection and reduce false and missed alarms. Compared to traditional algorithms that require manual design of features, deep learning can automatically extract key features from data, reducing the need for human intervention. Moreover, deep learning models can be trained in different environments, enabling them to adapt to a wide range of situations such as changes in light, background interference, etc., which enables real-time smoke detection and is suitable for surveillance and warning systems. Therefore, deep learning-based smoke detection techniques for video images are increasingly widely used. This innovative smoke detection technology is adept at navigating complex and dynamic environments, where sensitivity is paramount. It not only addresses the longstanding challenges associated with poor real-time performance, low accuracy, and high false detection rates, but also significantly enhances them.

The fundamental research of deep learning-based smoke detection is the object detection algorithm. In general, object detection algorithms are mainly classified into two categories: two-stage [4] and one-stage [5]. The two-stage algorithm firstly generates candidate regions, then classifies and regresses the candidate region, while the one-stage algorithm treats the target detection problem as a single regression issue and directly outputs the category and location information of the target. No pre-selected boxes need to be generated, allowing for direct extraction of features from the network to predict both the classification and location of the object. Therefore, the one-stage algorithm is characterized by fast speed and good real-time performance. This paper introduces a pioneering fire smoke detection algorithm leveraging a one-stage architecture.

Technically speaking, the main contributions of the paper can be summarized as follows:

(1) We firstly propose a novel DCNFYOLO network based on duel convolution neural network and YOLO series for smoke detection in complex environments. Numerical experiments of qualitative and quantitative comparisons with state-of-the-art methods demonstrate the superiority of the proposed method, which achieves a desirable precision in smoke detection.

(2) The DCNFYOLO method reduces the model size and computational burden. It is superior to current methods in detection speed and has a high efficiency by integrating the ECA attention mechanism.

(3) The Wise-IOU loss function is designed and assigns the corresponding gradient gain for different outliers to improve the regression precision. The proposed DCNFYOLO is tested in both our dataset and public datasets, which have proved that the proposed network reduces the anticipation errors.

## 2. Related Work

### 2.1. Vision-Based Smoke Analysis

From the visual aspect, smoke detection methods include smoke recognition [6], smoke detection [7], and smoke segmentation [8]. Previous scholars usually use traditional features such as color, texture, and shape to detect smoke [9]; these methods always show a low

detection rate, high leakage rate, and high false detection rate. Presently, smoke detection algorithms are broadly categorized into two main groups: conventional smoke detection algorithms and those founded on deep learning methodologies [10]. Conventional smoke detection methods mainly rely on traditional machine learning classifiers by framing the smoke candidate region, followed by extraction of smoke features, and finally classification and detection [11]. The traditional algorithm defects lead to the inability to detect smoke targets in time and poor real-time detection of fire smoke. Starting from [12], deep learning-based methods [13–15] have been applied to smoke detection and segmentation, which are more promising solutions for smoke analysis.

## 2.2. Deep Learning-Based Smoke Detection

Fire detection in the early days was mainly realized through the Internet of Things (IoT) [16–18] and edge computing devices [19–22]. Recently, smoke detection utilizing deep learning has emerged as a popular and increasingly favored approach. Smoke detection based on deep learning can be segmented into three components: classification [23], detection, and segmentation [24]. Based on the current study, deep learning-based smoke detection algorithms are generally better than traditional algorithms in various aspects such as real-time detection and stability.

Recently, there have been intense research interests on deep learning algorithms to monitor smoke. Based on the MobileNetV2 network architecture in foggy surveillance environments, the Edge Intelligence-Assisted algorithm was proposed for smoke detection in industrial scenarios [25]. EdgeFireSmoke [26] takes full account of the detection time as well as GPU memory consumption, and it is a new lightweight CNN. Compared with Edge Intelligence-Assisted and Energy-Efficient [27] algorithms, EdgeFireSmoke exhibits negligible performance degradation while substantially reducing both the parameter count and the model dimensions, which can make it well-suited for aerial image analysis in video surveillance systems, especially when integrated with edge computing devices for CNN image processing. Feng et al. [28] integrated both static and dynamic features of smoke and used clustering and outer-circle detection to finalize the smoke detection. Mozaffari et al. [29] proposed an enhanced fire smoke detection model based on detection transformers (DETR). This model builds upon the DETR framework and incorporates deformable convolution along with the transformer relational modeling capabilities to achieve optimal sparse spatial sampling for detecting smoke. Li et al. [30] applied fusion-based deep network alignment to the VGG16 and ResNet-50 models to enhance the feature extraction capability of the networks. However, the method increases the computational complexity of the models while improving their detection accuracy.

## 2.3. Attention-Based Mechanism for Object Detection

Attention mechanisms in computer vision automatically select and emphasize the most relevant features or areas of input data, thereby reducing attention to irrelevant information. At the same time, it reduces the waste of computational resources, makes the model more efficient in processing complex images, and can combine multi-scale features to enhance the flexibility and accuracy of processing objects of different sizes.

To assess, detect, and classify the bottom of the ocean floor using sonar, Polap et al. [31] proposed a multi-attention model that focused on local and global sonar functions and fused them. This approach increases the amount of data and improves the effectiveness of the solution based on the digital twin layer of the generation and classification network model. In order to achieve accurate detection of ripe strawberries and their occlusion degree, Du et al. [32] introduced a Shuffle Attention (SA) mechanism at the end of the backbone network to enhance the extraction of effective information and improve the detection accuracy. To solve the problems of low contrast of infrared ship images, uneven distribution of ship size, and slow detection speed of unmanned ships, Wu et al. [33] introduced an effective multi-scale attention mechanism (EMA) based on YOLOv8 to improve the model's ability to detect objects at multiple scales and the speed of model inference.

Attention mechanisms are widely used in image classification, object detection, image segmentation, and other tasks, and have also made remarkable achievements in the field of natural language processing. By assigning weights to different features or regions, the attention mechanism improves the model understanding of importance, which in turn improves accuracy.

#### 2.4. YOLO-Based Object Detection

Recognition and localization are two main tasks of smoke detection. The essence of smoke recognition lies in extracting and categorizing smoke features within an image, while the primary focus of smoke localization is pinpointing and providing markers for regions indicative of smoke presence in an image [34].

Wang et al. [35] used YOLOv5 to provide a multi-object detection network for foggy driving scenarios. A new Feature Enhancement Module (FEM) was developed, integrating an attention mechanism to aid the detection network in allocating greater focus to the pertinent features within foggy scenes. This advancement ensures that the network prioritizes crucial information, thereby enhancing its performance in challenging weather conditions. Tu et al. [36], based on the YOLOv5 detection model, proposed a MoCo network structure that enhances the differentiation between targets and backgrounds by addressing features such as deformation, occlusion, lighting, and small objects. Yang et al. [37] introduced C3Ghost and Ghost modules into the YOLOv5s network architecture, which compress and reduce the model computational cost while maintaining detection accuracy and speed. Al-Smadi et al. [38] implemented a Stochastic Gradient Descent (SGD) optimizer to minimize detection time in the YOLOv5 object detection algorithm, while also fine-tuning the remaining fundamental parameters of the detection process. Comparisons with models such as Faster R-CNN and YOLOv4 showed that their approach resulted in a significant improvement in the model mAP value. Li et al. [39] developed a novel detector called YSCRM using YOLOv5 feature fusion layer, SK-Net, transformer encoder module, sparse training method, and various training techniques, which showed excellent performance in recognizing complex satellite components. Song et al. [40] proposed MS-YOLO, an innovative deep learning object detection framework that integrates multiple data sources, specifically combining millimeter-wave radar and vision. This novel approach aims to enhance detection capability through the synergistic effect of these complementary modalities.

In conclusion, while the methods mentioned above contribute to enhancing the accuracy of smoke detection, there remains a notable gap in research regarding the enhancement of model complexity and portability. Motivated by these facts, a novel DCNFYOLO network for smoke detection is proposed in this paper. The integration of efficient convolution, attention mechanisms, and a dynamic loss function has substantially reduced the computational complexity of the model. Additionally, these innovations facilitate rapid model convergence and optimize the utilization of loss resources.

### 3. Materials and Methods

#### 3.1. Dataset Acquisition and Preprocessing

The datasets used in this paper consist of two main parts, as shown in Table 1. The first part includes 6863 fire smoke images and 2418 non-smoke images obtained via the Internet. It contains different types of fire smoke scenarios including towns, factories, and forests, indoors and outdoors. The latter part comprises a collection of public datasets extensively utilized in the relevant literature, including the following: (1) the CVPR Lab-KMU Fire and Smoke Database; (2) the Smoke dataset sourced from the State Key Laboratory of Fire Science, University of Science and Technology of China. The typical samples of the datasets are shown in Figure 1.

Table 1. Experimental datasets.

| Types               | Dataset        | Smoke Image | Non-Smoke Images | Total  |
|---------------------|----------------|-------------|------------------|--------|
| Self-built datasets | Training set   | 4200        | 1050             | 5250   |
|                     | Test set_1     | 1250        | 684              | 1934   |
|                     | Validation set | 1413        | 684              | 2097   |
|                     | Total          | 6863        | 2418             | 9281   |
| Public datasets     | Test set_2     | 1300        | 400              | 1700   |
|                     | Test set_3     | 1300        | 400              | 1700   |
|                     | Total          | 2600        | 800              | 3400   |
| All                 | Total          | 9463        | 3218             | 12,681 |

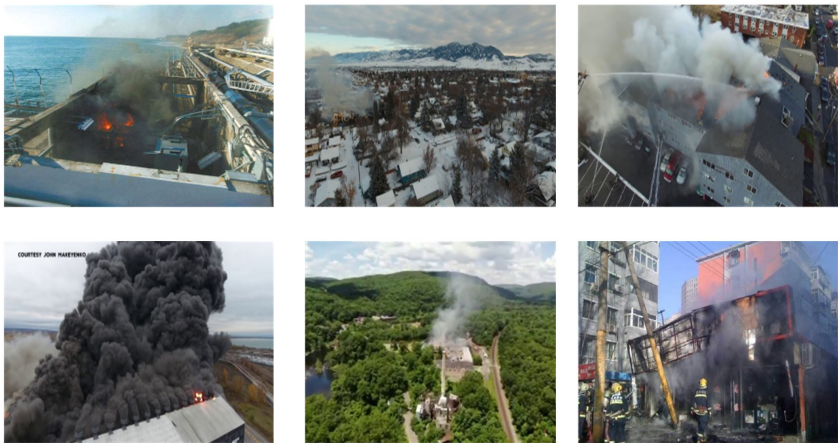


Figure 1. Typical examples of the public datasets.

Images Preprocessing

To improve the robustness of the model, avoid overfitting, and enhance the generalization ability of the model, data enhancement methods are used in this paper, including mosaicing, adaptive image filling, random cropping, and flipping. The model input is illustrated in Figure 2.

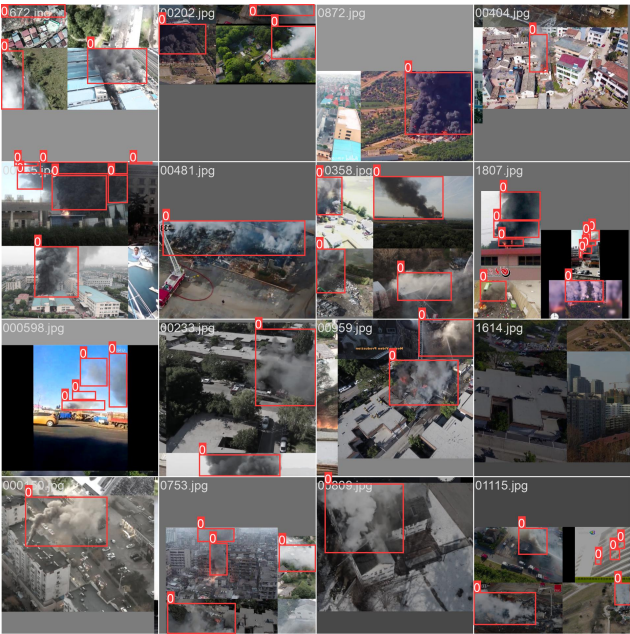
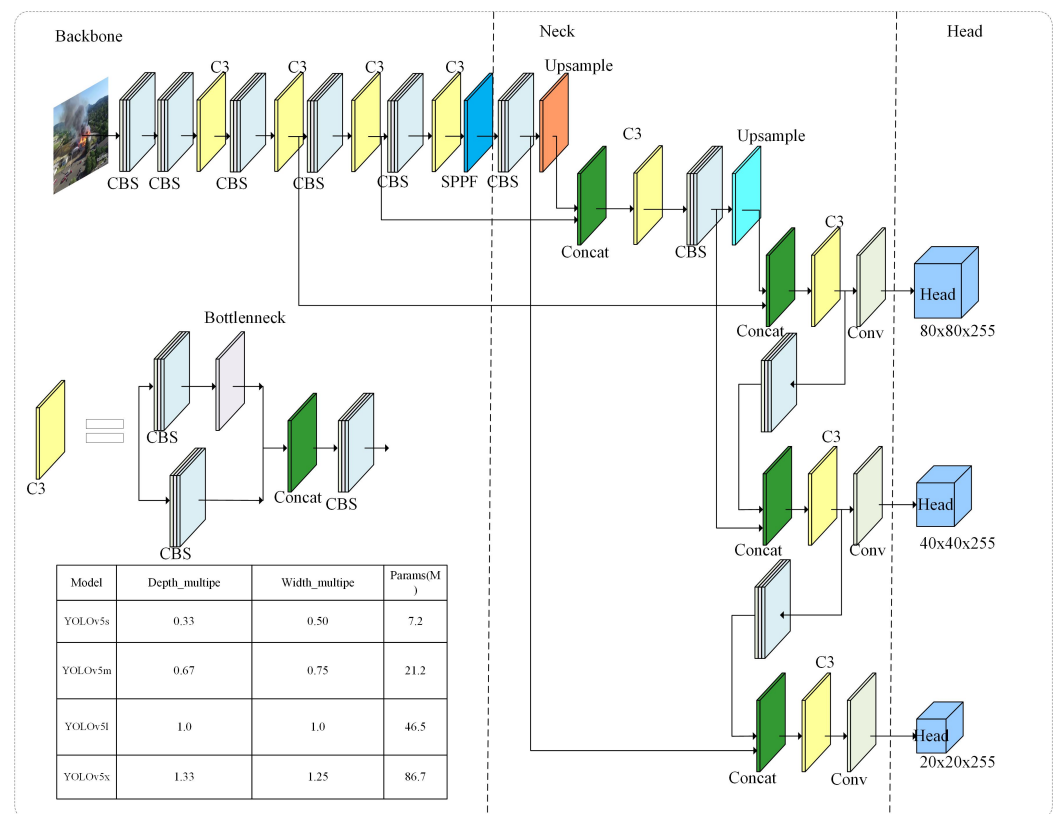


Figure 2. Data-enhanced image.

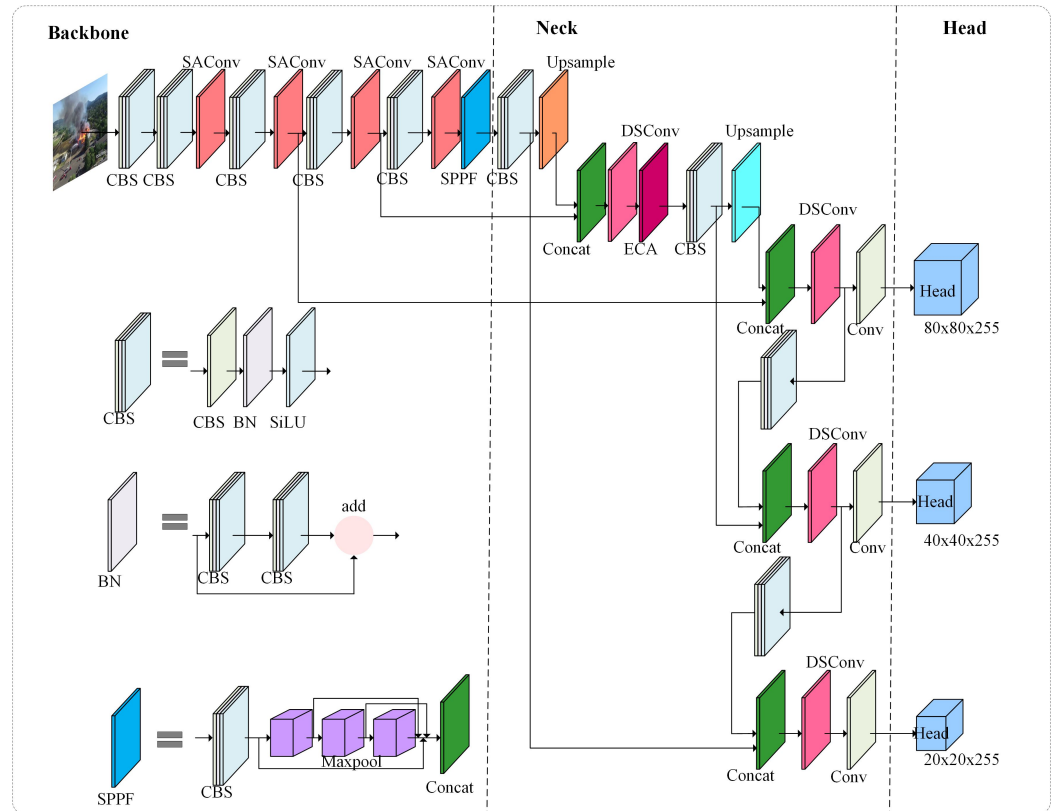
### 3.2. Proposed Method

The latest YOLOv5 version 7.0 is used in this paper. The network structure is shown in Figure 3. There are three main components of the network structure: head structure, neck structure, and backbone network. Features are extracted and the feature map undergoes progressive reduction through the backbone network. The main structure contains a ConvBNSiLU(CBS) down-sampling processing module, a C3 module, and Spatial Pyramid Pooling-Fast (SPPF) module. The neck utilizes the Feature Pyramid Network (FPN) to merge shallow graphical features with deeper semantic features, aiming to acquire more comprehensive features. The head layer functions as the detect module and comprises three convolutions corresponding to three feature detection layers, facilitating upscaling or downscaling as needed.



**Figure 3.** YOLOv5 network structure.

As shown in Figure 4, the proposed method in this paper builds a novel smoke detection model on a computer. The input smoke image first passes through multiple CBS and SAConv layers in the backbone network to extract deeper features. Towards the end of the backbone, the faster SPPF module is employed to combine multi-scale features. In this process, feature representations from different scales of the same feature map are seamlessly integrated, thereby improving the accuracy of feature extraction. Then, the feature maps are sent to the neck section, including ECA and DSConv lightweight network structure. ECA can take into account attention across both channel and spatial dimensions, allowing the model to hone in more precisely on relevant channel information by acquiring adaptive channel weights. Introducing DSConv offers the dual benefits of reducing model size and computational load, thereby accelerating the detection speed. Ultimately, the processed information is directed to the head section to finalize the entire target detection process.



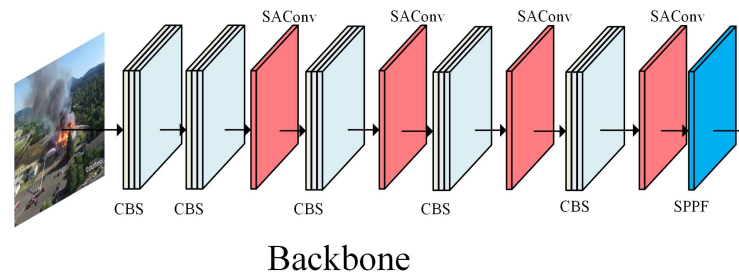
**Figure 4.** Network structure diagram of DCNFYOLO.

### 3.2.1. SAC-Based Backbone Network for DCNFYOLO

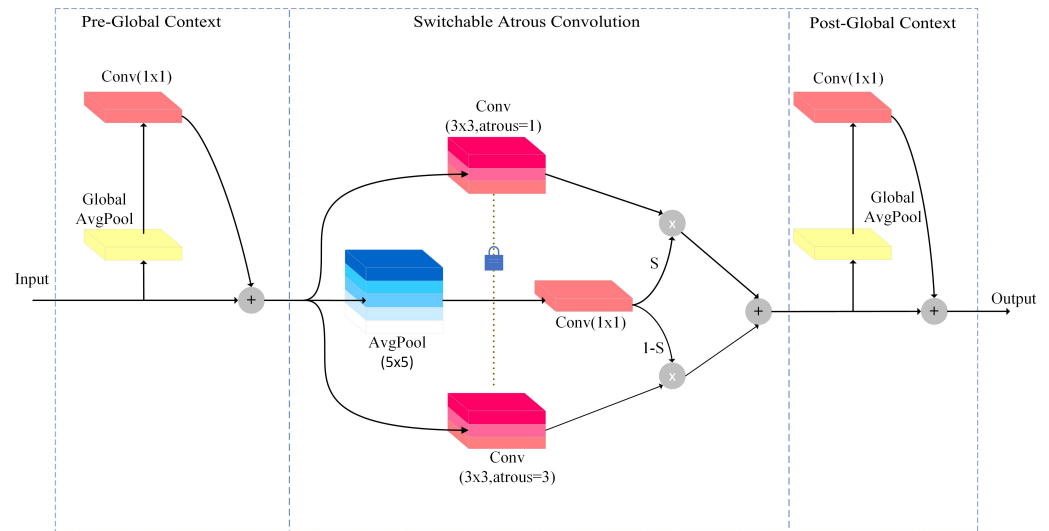
To enhance smoke detection accuracy, this paper substitutes the C3 module in the backbone network with the SAConv module. This change boosts the network's ability to extract smoke features while minimizing redundancy in the extracted data, enabling deeper feature extraction without increasing inference time. Figure 5 illustrates the layout of the DCNFYOLO backbone network, where the first two CBS modules are designed to reduce the input image size while increasing the number of image channels. The SAConv module consists of three main components: two global context modules and one SAC component. The network structure is illustrated in Figure 6. The smoke dataset is output by the CBS module and transferred to the SAC network for further feature extraction. SAConv can convolve the same input features with different Atrous rates and collect the results using a switching function. The switching function is spatially correlated, and each position of the feature map can be controlled by a different switch to control the output of the SAC, thus enhancing the feature extraction of the neural network on the smoke dataset. This converts the convolutional layer to SAC, as shown in Equation (1).

$$y = S(x) * \text{Conv}(x, w, 1) + (1 - S(x)) * \text{Conv}(x, w + \Delta w, r) \quad (1)$$

where  $x$  and  $y$  are the input and output of the switchable atrous convolution, and  $r$  is a hyperparameter of the SAC representing the atrous rate,  $w$  represents the weights, and  $\text{Conv}(x, w, r)$  denotes the convolution operation. The switch function  $S(x)$  is an average pooling layer of  $5 \times 5$  kernels followed by the  $1 \times 1$  convolutional layer. The insertion of a lightweight global context module before and after the main component SAC makes the judgment of the switching function  $S(x)$  more stable.



**Figure 5.** SAC-based backbone network.



**Figure 6.** SAC network structure diagram.

### 3.2.2. ECA Attention Mechanism Module

Smoke object detection feature extraction requires datasets with high pixel resolution. However, when smoke cannot be distinguished from complex backgrounds at lower resolutions or when the smoke is not distinct, using the original YOLOv5 network for smoke feature extraction may result in the loss of smoke feature information. To address this issue, this paper integrates the ECA module into the neck of the DCNFIYOLO network structure. The ECA module is derived from the Squeeze and Excite Network (SE-Net) module. The SE-Net architecture is depicted in Figure 7. The ECA mechanism employs local convolutions to capture inter-channel relationships rather than relying on global information compression. This method mitigates the potential information loss associated with the fully connected layers in the SE mechanism, thereby preserving feature information more effectively. The working process of the ECA mechanism can be illustrated in Figure 8. The ECA mechanism module integrates a  $1 \times 1$  convolutional layer directly following the global average pooling layer, eliminating the need for a fully connected layer. This approach circumvents dimensionality reduction and effectively captures cross-channel interactions. Moreover, utilizing one-dimensional convolution for effective implementation of local cross-channel interactions and extraction of inter-channel dependencies, the ECA mechanism attains superior outcomes with reduced parameters. This substantially decreases the complexity of the model while upholding its performance. The convolutional kernel size can be adaptively varied by a function that allows more cross-channel interactions for layers with larger channel counts.

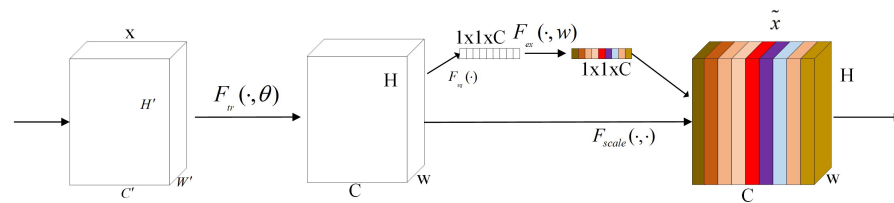


Figure 7. SE attention structure diagram.

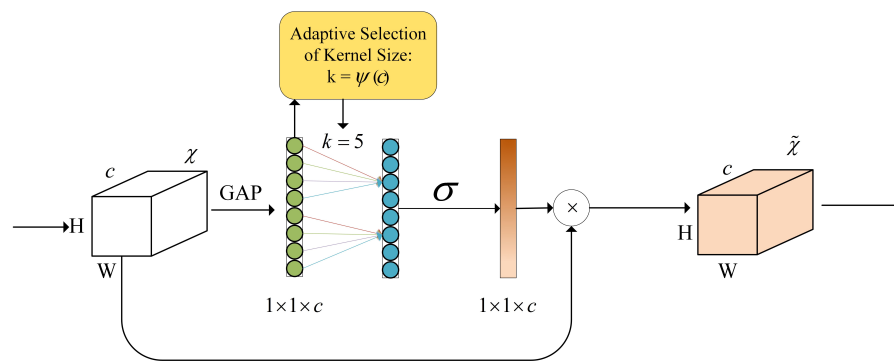


Figure 8. ECA structure diagram.

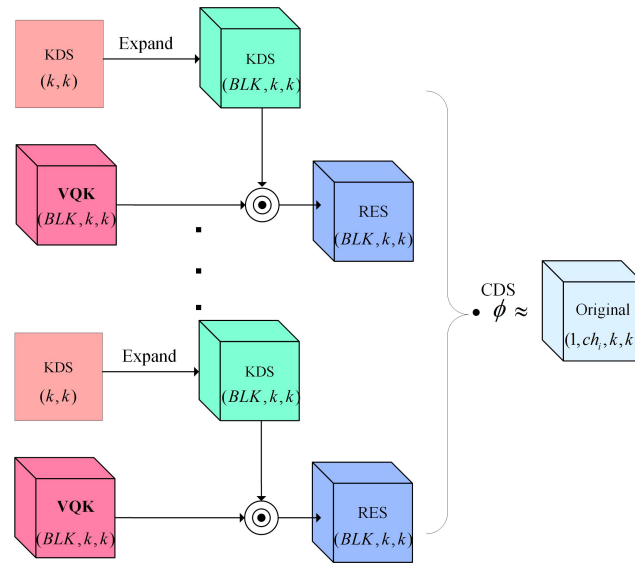
The formula for dynamically determining the size of the convolution kernel is provided in Equation (2).  $c$  denotes the number of channels, and the symbols  $\gamma$  and  $b$  are employed to adjust the ratio between the size of the convolutional kernel and the number of channels  $c$ . In the experiments,  $\gamma$  and  $b$  are set to 2 and 1, respectively.

$$k = \left\lceil \frac{\log_2(c)}{\gamma} + \frac{b}{\gamma} \right\rceil \quad (2)$$

The ECA attention mechanism handles local dependencies between channels. By using one-dimensional convolution operations, the ECA mechanism captures local dependencies between adjacent channels rather than global channel interactions, ensuring a more efficient smoke detection model. In addition, ECA extracts the global features of each channel through global average pooling, representing the channel response intensity to the entire image, and extracts the channel importance weights. Through the local convolution of channel features, ECA generates a set of attention weights, reflecting the importance of each channel to task objectives.

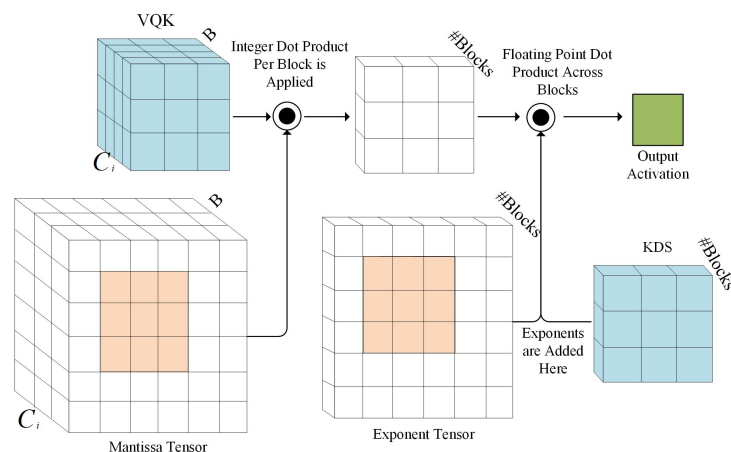
### 3.2.3. Distribution Shifts Convolution

To facilitate the deployment of the model, this paper introduces DSConv efficient convolution in the neck network structure to increase the detection speed and reduce the computational cost of the model. DSConv decomposes a traditional convolutional kernel into two parts: a variable quantization kernel (VQK) and a kernel distributed shift (KDS). Reduced memory consumption and increased processing speed can be attained through the exclusive storage of integer values in the VQK, while preserving identical output compared to the initial convolution kernel through the utilization of kernel-based and channel-based distributed shifts. DSConv simulates the behavior of a convolutional layer by using quantization and distribution offsets. The specific setup is illustrated in Figure 9. The BLK hyperparameter indicates the block size. The original convolution tensor size is  $(cho, chi, k, k)$ , where  $cho$  is the number of channels in the next layer,  $chi$  is the number of channels in the current layer, and  $k$  is the width and height of the kernel. The above settings are performed for each channel and each block. DSConv reduces the computational cost of the model with VQK and KDS components.



**Figure 9.** This process is performed for each channel and each block.

The specific data path is shown in Figure 10. Initially, each block of the VQK and mantissa tensor undergoes dot product operations, yielding individual values. All computations are executed using low-bit fixed-point operations, significantly reducing processing time. Following block multiplication, a tensor of the same size as the exponential tensor and KDS is derived. Subsequently, the exponent tensor is integrated with the KDS tensor by adding their respective values, resulting in a tensor of floating-point numbers. Finally, this tensor is multiplied with the outcome of the VQK and KDS product to generate a floating-point number as the output activation. Throughout this process, inference remains highly parallelizable, akin to standard convolution, where the majority of multiplications are performed using integer rather than floating-point arithmetic. This optimization conserves energy, bandwidth, and computation time.



**Figure 10.** Example of convolution being performed.

### 3.2.4. WIoU Loss Function

YOLOv5 uses the CIoU loss function to calculate the loss in the network. CIoU introduces the distance between the predicted bounding box and the center point of the real bounding box based on DIoU, which makes the loss function focus more on the position of the bounding box. However, CIoU has a high complexity and does not consider the balance between hard complex copies and simple samples. This could exacerbate the punishment for low-quality instances, potentially leading to a decline in the model generalization performance. To address these concerns, the WIoU loss function is utilized to calculate the

loss within the architecture of the DCNFYOLO network. Both geometric factors, distance and aspect ratio, have to be considered, since the training data inevitably contain low-quality examples. The WIoU loss function can weaken the influence of geometric factors when the anchor box overlaps well with the target box, and less intervention in training enables the model to obtain a better generalization ability.

The minimum bounding box formed by the real and predicted boxes is shown in Figure 11. The area calculation of the green boundary box and red boundary box splicing set is shown in Equation (3). The  $IoU$  is computed as illustrated in Equation (4).

$$S_u = wh + w_{gt}h_{gt} - W_iH_i \quad (3)$$

$$\mathcal{L}_{IoU} = 1 - IoU = 1 - \frac{W_iH_i}{S_u} \quad (4)$$

There are three variations of WIoU: WIoU v1, WIoU v2, and WIoU v3. To enhance the model convergence speed and stability, this study utilizes the WIoU v3 loss function. WIoU v3 is derived by introducing non-monotonic focusing coefficients to WIoU v1. WIoU v1 can be calculated from Equation (5). The specific value of  $\mathcal{R}_{WIoU}$  is shown in Equation (6).

$$\mathcal{L}_{WIoUv1} = \mathcal{R}_{WIoU} \mathcal{L}_{IoU} \quad (5)$$

$$\mathcal{R}_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (6)$$

where  $\mathcal{R}_{WIoU} \in [1, e)$ ,  $\mathcal{L}_{IoU} \in [0, 1]$ ; when the anchor frame coincides well with the target, the focal length between the high-quality anchor frame ground  $\mathcal{R}_{WIoU}$  and its center point can be significantly reduced.  $W_g$  and  $H_g$  are the dimensions of the smallest closed frame, and the superscript \* refers to the separation of  $W_g$  and  $H_g$  from the computational map to prevent the generation of  $\mathcal{R}_{WIoU}$  gradients that prevent convergence.

The WIoU loss function evaluates the quality of the anchor frame by calculating the outliers of the anchor frame, which is defined as  $\beta = \frac{(1-IoU)^*}{1-IoU} \in [0, +\infty)^{(*)}$ , and assigns different gradient gains accordingly. In order to make real-time decisions that best fit the current gradient gain allocation strategy,  $\mathcal{L}_{WIoUv3}$  can be introduced. The WIoU loss function continuously determines the most appropriate gradient gain assignment strategy in the current situation.  $\mathcal{L}_{WIoUv3}$  can be calculated by Equation (7).

$$\mathcal{L}_{WIoUv3} = r \mathcal{L}_{WIoUv1} \quad (7)$$

$$r = \frac{\beta}{\delta \alpha^{\beta-\delta}} \quad (8)$$

In Equation (8),  $r$  is the non-monotonic focusing coefficient. Experiments verify that bounding box localization and smoke prediction are more effective when  $\alpha$  and  $\delta$  are set to 1.4 and 2.6, respectively. The WIoU algorithm assigns varying gradient gains to different outliers, enabling the training model to dynamically allocate gradient gains based on the current situation. This adaptive strategy promotes smoother convergence, enhances regression accuracy, and reduces prediction errors.

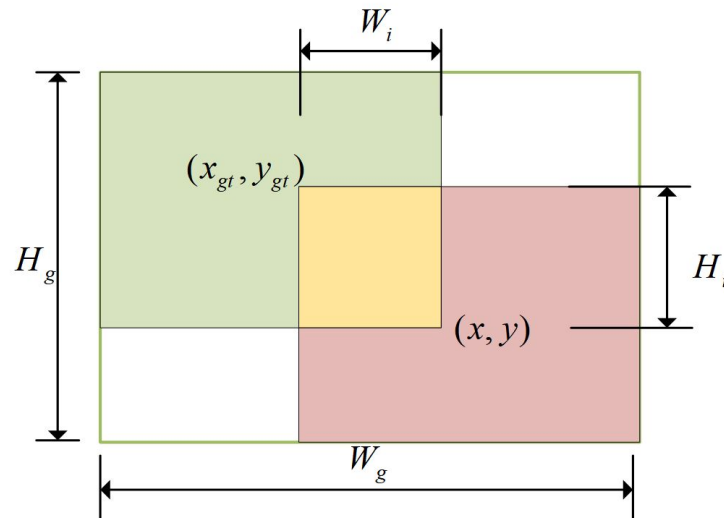


Figure 11. Minimum bounding box formed by the true and predicted boxes.

## 4. Results

### 4.1. Experimental Environments

This experiment uses PyTorch version 1.10.0, Python version 3.8, and Ubuntu (20.04) system, as well as an NVIDIA RTX 3090 graphics card with 24 GB of video memory, Cuda version 11.3, and the following CPU configuration: 12 vCPU Intel(R) Xeon(R) Platinum 8255C 2.50 GHz. The training of the model incorporates a multi-scale training approach, with the batch size set at 16 and spanning 100 iterations, while leveraging the SGD optimizer. The input image size is  $640 \times 640$ . The initial learning rate of the model is set to 0.01, while the SGD starting momentum starts from 0.096.

### 4.2. Model Evaluation Indicators

This experiment uses  $P$  (Precision),  $AP$  (Average Precision),  $R$  (Recall), and FPS (Frames per Second) to estimate the prediction results of the model. Furthermore, this investigation employs the size of the model FLOPs to evaluate its performance. FLOPs represent the quantity of floating-point operations utilized to compute the model complexity. The expressions of  $P$ ,  $R$ , and  $AP$  are derived as the following Equations (9), (10), and (11), respectively.

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$AP = \int_0^1 P dR \quad (11)$$

The variable  $TP$  represents the count of positive samples, indicating the number of bounding boxes where the model accurately predicts smoke.  $FP$  is the false positives, which is the number of bounding boxes where the model incorrectly detected smoke targets.  $FN$  is the number of smoke targets that the model failed to detect. The higher  $AP$  value correlates with the increased recognition accuracy of the algorithm.

### 4.3. Analysis of Methodology and Effectiveness

To evaluate the efficacy of the proposed model, this paper conducts ablation experiments. These experiments compare model parameters and FLOPs between YOLOv5 and the proposed model, while also comprehensively evaluating the impact of the chosen loss function, attentional mechanism, and network structure. YOLOv5 is utilized as the benchmark model for each experiment.

#### 4.3.1. Comparisons of Attention Mechanisms

To validate the effectiveness of the selected attention mechanisms, this paper incorporates SA (Spatial Attention), SE (Squeeze-and-Excitation) channel attention, CBAM (Convolutional Block Attention Module), and ECA into the model backbone for training and comparison. The experimental results are presented in Figure 12 and Table 2.

SA mainly focuses on spatial dimensions while neglecting the importance of channel dimensions, which results in a 1.1% decrease in model accuracy when combined with SA. The SE module enhances the network's ability to represent useful features, making the model more effective in selecting and extracting smoke features. However, incorporating SE into the network introduces additional fully connected layers and pooling operations, which increases the overall model complexity and can lead to reduced detection speed. CBAM leveraging spatial information through channel dimension reduction and subsequent convolution yields an improvement in detection accuracy. Nevertheless, its reliance on convolution limits its ability to capture broader relationships beyond local contexts. The ECA mechanism uses one-dimensional convolution to efficiently capture local cross-channel interactions, improving model performance while reducing complexity. The experimental results illustrate that the integration of the ECA mechanism into the backbone network improves the  $P$  of the model by 6.7% compared to the original network. Experiments verified the effectiveness of the attention mechanism selected in this paper.

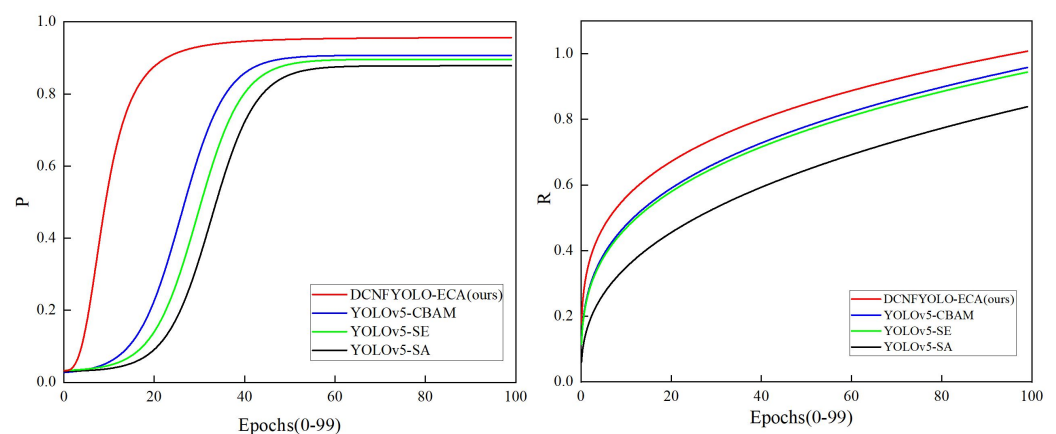


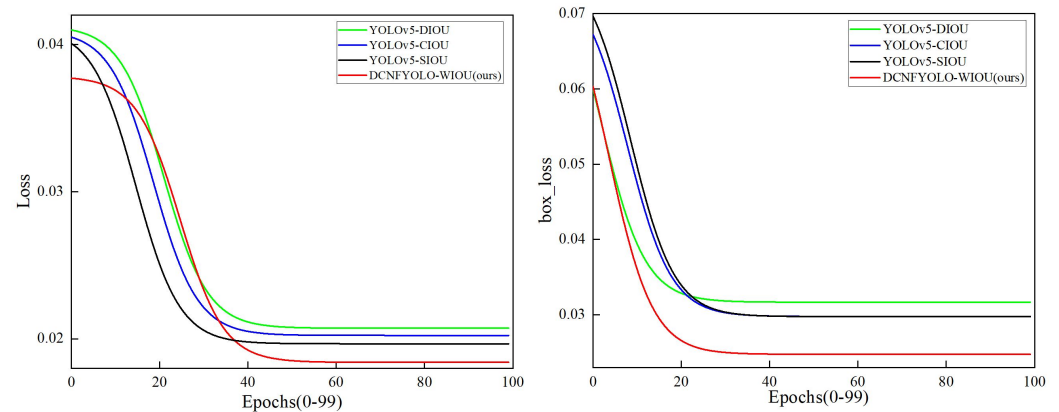
Figure 12. Comparisons of different attention mechanisms.

Table 2. Experimental results of different attentional mechanisms.

| Attention         | $P$   | $R$   | FPS  |
|-------------------|-------|-------|------|
| Backbone          | 0.889 | 0.896 | 73.4 |
| YOLOv5+SA         | 0.878 | 0.835 | 72.6 |
| YOLOv5+SE         | 0.896 | 0.943 | 69.2 |
| YOLOv5+CBAM       | 0.907 | 0.957 | 72.5 |
| YOLOv5+ECA (Ours) | 0.956 | 0.986 | 73.9 |

#### 4.3.2. Comparisons of Loss Functions

To validate the selection of the WIoU, this study evaluates the performance of different loss functions, including SIoU, DIoU, and WIoU. The comparison is made with the YOLOv5 CIoU loss function. Based on the analysis of Figure 13 and Table 3, the selected loss function in this research shows notable improvements in Box\_loss, overall Loss, and  $AP$  metrics compared to other alternatives. Specifically, both Box\_loss and overall loss are reduced by 0.5% and 0.184%, respectively, while  $AP$  sees a notable improvement of 0.9% when compared to the CIoU loss function utilized in YOLOv5.



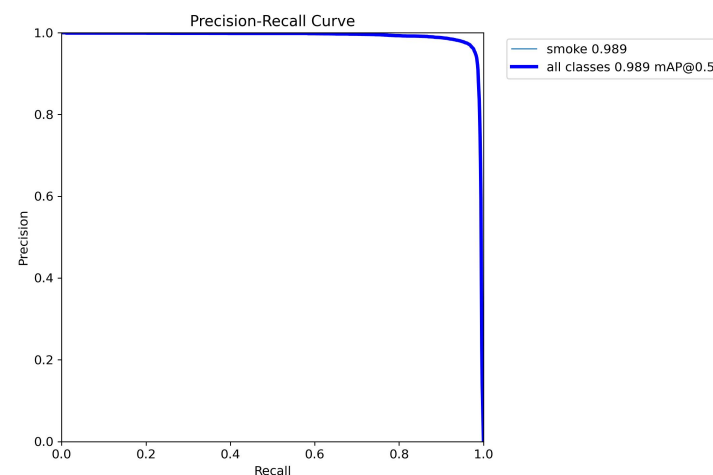
**Figure 13.** Comparisons of different loss functions.

**Table 3.** Performance of different loss functions.

| Loss Function      | Box_loss | Loss    | AP    |
|--------------------|----------|---------|-------|
| YOLOv5+CIoU        | 0.02975  | 0.02071 | 0.911 |
| YOLOv5+SIoU        | 0.02974  | 0.02023 | 0.920 |
| YOLOv5+DIoU        | 0.03167  | 0.01965 | 0.922 |
| YOLOv5+WIoU (Ours) | 0.02474  | 0.01839 | 0.931 |

#### 4.3.3. *P-R* Curve

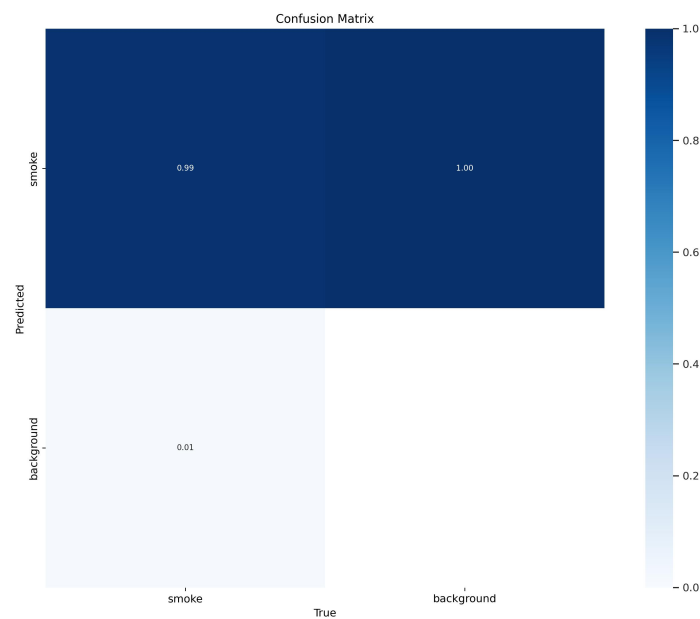
The *P-R* curve, as shown in Figure 14, has the recall rate on the horizontal axis and the precision on the vertical axis. The closer the curve is to the top right corner, the better the overall performance of the model. According to the *P-R* curve, the smoke detection method proposed in this paper has high accuracy.



**Figure 14.** *P-R* curve.

#### 4.3.4. Confusion Matrix for Dataset Construction

The training process in this paper incorporates a confusion matrix, and it is illustrated in Figure 15. Each column of the confusion matrix signifies the predicted category, with the total count indicating the number of data instances predicted within that category. On the other hand, each row corresponds to the attribute category of the data, with the total count showcasing the number of data instances belonging to that category.



**Figure 15.** Label distribution.

#### 4.4. Comparisons among Different Models

To validate the efficacy of the proposed model, experiments and simulations are conducted to compare its performance with that of prevalent target detection networks under the same conditions, using the same smoke dataset. Table 4 presents the experimental results of different models. The SSD is a one-stage detector that combines the generation of candidate boxes and classification into a single process, making the training process simple and efficient. However, when detecting small smoke targets on shallow feature maps, the resolution of these feature maps is relatively low, resulting in poorer detection performance for small smoke targets. The Faster R-CNN is a two-stage detector, which needs to generate candidate regions first and then perform classification and regression, so its detection speed is relatively slow and not applicable to real-time smoke detection. DeepSmoke effectively extracts the smoke features in the image by deep convolutional neural network, so it has high detection accuracy in the smoke detection task, but DeepSmoke has a large amount of computation and high hardware resource requirements, which makes it unsuitable for real-time processing or running on resource-constrained devices. The proposed DCNFOLO in this study demonstrates significant improvements in  $P$ ,  $R$ , and  $AP$  compared to the YOLOv4, YOLOv4-tiny, and YOLOv5 algorithms. Specifically, compared to the widely used YOLOv5 algorithm, DCNFOLO achieves improvements of 7.7%, 7.4%, and 7.7% in  $P$ ,  $R$ , and  $AP$ , respectively, further validating the effectiveness of this method.

**Table 4.** Test results of different detection networks.

| Model        | $P$   | $R$   | $AP$  | Model Size/MB |
|--------------|-------|-------|-------|---------------|
| SSD          | 0.836 | 0.832 | 0.841 | 142           |
| Faster R-CNN | 0.854 | 0.841 | 0.862 | -             |
| DeepSmoke    | 0.902 | 0.863 | 0.896 | -             |
| YOLOv4       | 0.905 | 0.901 | 0.931 | 244           |
| YOLOv4-tiny  | 0.899 | 0.879 | 0.892 | 22.6          |
| YOLOv5       | 0.889 | 0.896 | 0.912 | 14.1          |
| DCNFOLO      | 0.966 | 0.97  | 0.989 | 18.6          |

#### 4.5. Ablation Experiment

To further explore the impact of various enhanced network structures on detection algorithm performance, this paper uses YOLOv5 as a reference point and conducts quantitative experiments to validate the effectiveness of these enhancements. The detection part includes  $P$ ,  $R$ ,  $AP$ ,  $AP_{0.5:0.95}$ , and GFLOPs. Ablation experiments are conducted using a self constructed dataset, and the results are summarized in Table 5. The table demonstrates the impact of various enhancements on detection performance when integrated into the YOLOv5 architecture. The combination of SConv with the original YOLOv5 backbone network increases the  $AP$  by 1.3%. Introducing the ECA mechanism in the neck part leads to a substantial increase in  $AP$  by 2.8%. Additionally, the incorporation of  $WIoU$  improves  $AP$  by 4.8%. Furthermore, integrating DConv results in a reduction in computation by 2.7 GFLOPs. When all these enhancement strategies are implemented simultaneously in the YOLOv5 model, significant improvements are observed. The  $AP$  value increases by 7.7% compared to the original pre-training weights. Moreover,  $AP_{0.5:0.95}$  increases by 2.45%,  $P$  increases by 7.7%, and  $R$  increases by 7.4%, with only a marginal increase of 0.1 GFLOPs in computation. These results demonstrate that the DCNFIYOLO network shows significant improvements compared to the original network.

Table 5. Ablation experiments.

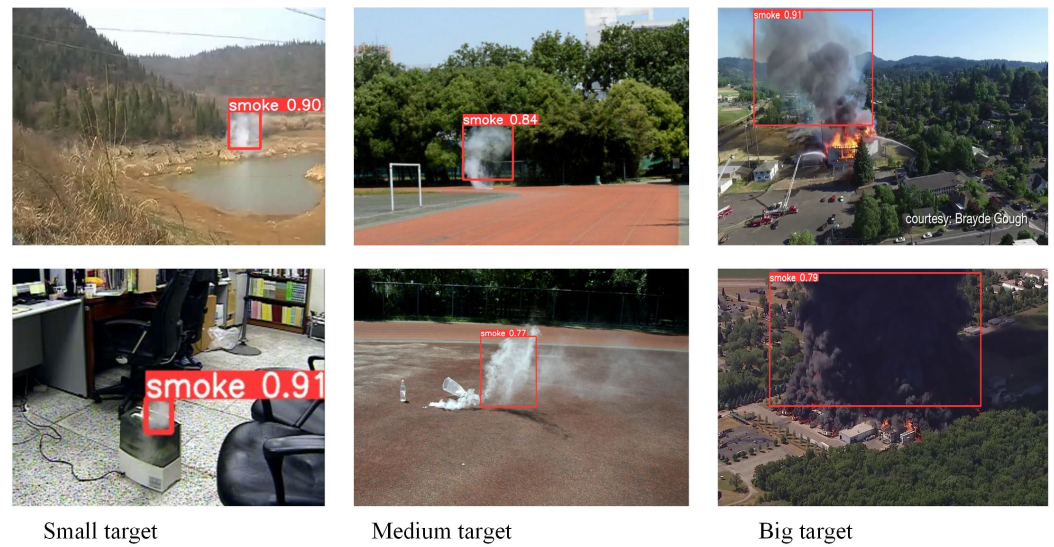
| Number            | SConv | DConv | ECA | WIoU | Precision | Recall | AP    | AP <sub>0.5:0.95</sub> | GFLOPs |
|-------------------|-------|-------|-----|------|-----------|--------|-------|------------------------|--------|
| YOLOv5            |       |       |     |      | 0.889     | 0.896  | 0.912 | 0.50                   | 15.9   |
| YOLOv5+SConv      | ✓     |       |     |      | 0.915     | 0.913  | 0.925 | 0.53                   | 19.5   |
| YOLOv5+DConv      |       | ✓     |     |      | 0.896     | 0.884  | 0.900 | 0.48                   | 13.2   |
| YOLOv5+ECA        |       |       | ✓   |      | 0.918     | 0.915  | 0.940 | 0.55                   | 16.6   |
| YOLOv5+WIoU       |       |       |     | ✓    | 0.923     | 0.927  | 0.960 | 0.62                   | 15.9   |
| YOLOv5+SConv+WIoU | ✓     |       |     | ✓    | 0.943     | 0.932  | 0.944 | 0.57                   | 19.6   |
| YOLOv5+DConv+WIoU |       | ✓     |     | ✓    | 0.924     | 0.917  | 0.940 | 0.59                   | 13.4   |
| YOLOv5+ECA+WIoU   |       |       | ✓   | ✓    | 0.947     | 0.943  | 0.964 | 0.70                   | 17.3   |
| DCNFIYOLO         | ✓     | ✓     | ✓   | ✓    | 0.966     | 0.970  | 0.989 | 0.75                   | 16.0   |

#### 4.6. Analysis of Visualization Results

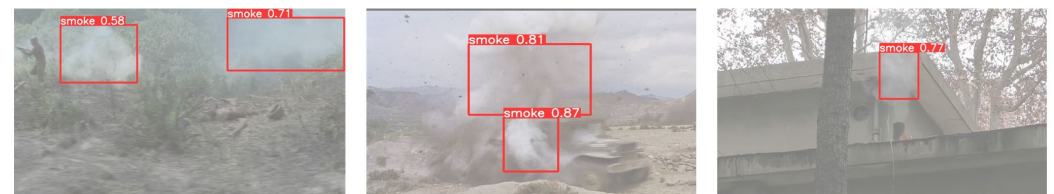
This study evaluates the DCNFIYOLO model's ability to maintain high accuracy and detection rates in challenging scenarios, focusing on smoke imagery from various environments. By analyzing these images, the study seeks to validate the model robustness and accuracy in detecting smoke objects amidst dynamic and varied environmental factors.

##### 4.6.1. Smoke Detection Results for Different Category Backgrounds

The size of smoke can be categorized into three types: small, medium, and large targets, with each type further divided into three common smoke colors: white, gray, and black. The test results depicted in Figure 16 reveal that the detection accuracy of smoke surpasses 90% across various complex scenes. Notably, the detection rate of small targets exceeds that of medium and large targets. This phenomenon can be attributed to the integration of SConv and the ECA mechanism into the network structure. These improvements facilitate improved feature extraction for small smoke targets during training, thereby strengthening the model's ability to detect them. Considering that small smoke targets typically emerge in the early stages of a fire, the enhanced model demonstrates increased versatility across different application scenarios. Furthermore, the study also evaluates the performance of smoke detection in disrupted environments, taking into account both the detection rate and accuracy. Illustrated in Figure 17, the self-constructed dataset is utilized to introduce various disruptive elements such as blur, fog, and background clutter, among others, into the original images for detection purposes. Remarkably, the DCNFIYOLO model achieves a detection rate exceeding 90% even under strong interferences. These test results affirm the model's robustness and capability to maintain high detection performance despite challenging environmental conditions.



**Figure 16.** Detection results of smoke targets of different sizes.



**Figure 17.** Smoke target detection in complex backgrounds.

In addition, the experimental scheme includes the CVPR Lab-KMU Fire and Smoke Database and the smoke dataset from the State Key Laboratory of Fire Science, University of Science and Technology of China, as shown in Figure 18. The confidence is closely related to the detection model and the image resolution. With respect to the public datasets, high detection accuracy can be achieved. Therefore, the superiority of the proposed smoke detection algorithm in this paper is further validated.



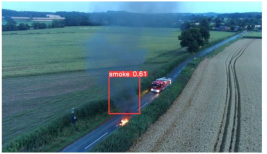
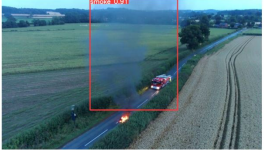
**Figure 18.** Testing on public datasets.

#### 4.6.2. Comparisons of Detection Results among Different Models

To further validate the detection capabilities of the enhanced model, this study conducts a comprehensive comparison with SSD, Faster R-CNN, and YOLOv5, all trained for the same number of epochs (100). The detection results are summarized in Table 6. Of note is that the SSD detection model mistook the trees in group (b) as smoke targets, and the SSD model has poor detection capability, high leakage rate, and relatively low confidence in detecting smoke. For YOLOv5 and Faster R-CNN algorithms, the confidence

level is comparatively low and fails to detect all the smoke, and the leakage detection rate is high. The DCNFYOLO algorithm enhances the extraction of smoke features in the network structure, which has a higher confidence level for smoke detection of small targets.

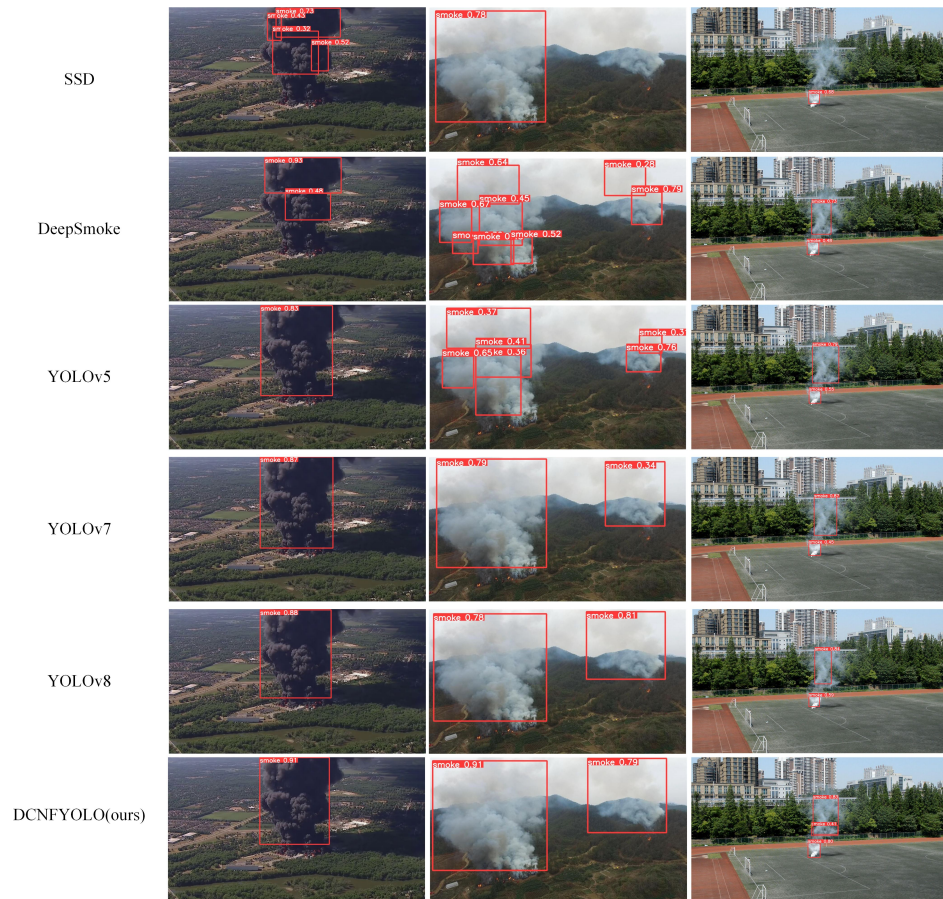
**Table 6.** Comparisons of experimental results among different detection models.

| Method       | Detection Result                                                                   |                                                                                     |                                                                                      |
|--------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| SSD          |   |   |   |
| Faster R-CNN |   |   |   |
| YOLOv5       |   |   |   |
| Ours         |  |  |  |
|              | (a)                                                                                | (b)                                                                                 | (c)                                                                                  |

#### 4.6.3. Comparisons of Advanced Model Detection Results on Common Datasets

To further validate the superiority of the proposed smoke detection model, the experimental part includes a comparison with the current state-of-the-art model, as shown in Figure 19. The SSD model exhibits low accuracy in detecting small objects due to its reliance on fixed default frames, which may not be able to accommodate the diversity and complexity of small smoke targets, leading to missed detections. The performance of the DeepSmoke model depends heavily on the quality and diversity of the training data, and the lack of diversity can lead to insufficient robustness to unknown smoke patterns. In addition, its processing speed may not be able to meet the real-time detection requirements in complex scenes.

Furthermore, this study introduces a comparison between different YOLO family models, keeping consistent preset training parameters and setting the training to 100 calendar hours. The detection results of various YOLO models are shown in Figure 19. It can be seen that the detection results of YOLOv5, YOLOv7, and YOLOv8 show low confidence and are therefore not suitable for outdoor fire smoke scenarios. In contrast, the confidence and detection rates of the models proposed in this paper are higher than those of the above models, and there is no leakage or false detection, which further validates the effectiveness of the algorithms proposed in this study.



**Figure 19.** Testing on common datasets.

## 5. Conclusions

This paper proposes a dual-convolution network and feature fusion approach aimed at achieving high-precision smoke detection. The multi-scenario smoke datasets were created to address the variability of smoke detection scenarios. To improve the detection of small targets, we incorporate SAConv into the backbone network, substantially boosting CNN performance in smoke target identification. Additionally, DSConv effectively diminishes the computational cost of the model, thereby enhancing deployment feasibility. Moreover, integrating the ECA mechanism and WIoU loss function ensures more stable convergence during training, further refining the model's efficacy. Experimental results show that the *AP* of the DCNFYOLO model increases by 7.7% when tested on public datasets. The precision, recall, and average detection speed reach 96.6%, 97%, and 70.9 FPS, respectively. Compared with current smoke detection algorithms, the proposed method shows a significant improvement in both detection precision and speed. However, due to the limitations of the dataset, deep learning methods cannot cover all complex scenarios and may miss detection in some scenarios (e.g., rainy and windy days). In the future, the generality of the algorithm can be further improved by refining the smoke datasets in different complex environments.

**Author Contributions:** Conceptualization, X.C., X.L. and Y.Z.; methodology, X.C. and X.L.; software, X.L.; validation, X.C., X.L. and B.L.; formal analysis, X.C. and X.L.; investigation, X.C., X.L. and B.L.; resources, X.C. and Y.Z.; data curation, X.C. and X.L.; writing—original draft preparation, Y.Z. and B.L.; writing—review and editing, X.L. and X.C.; visualization, Y.Z.; supervision, B.L.; project administration, X.C.; funding acquisition, X.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Nature Science Foundation of China under Grant No. 62176204, the Key Research and Development Program of Shaanxi Provincial Science and

Technology under Grant No. 2024-YBXM-052, the Scientific Research Project of Shaanxi Provincial Education Department under Grant No. 23JC031 and the Science and Technology Program of Xi'an, China under Grant No. 23GXFW0027.

**Data Availability Statement:** All experimental data in this paper are available upon request from the corresponding author.

**Conflicts of Interest:** Author Bing Liu was employed by the company Shaanxi Architectural Design Research Institute Co., LTD. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Abbreviations

|          |                                                          |
|----------|----------------------------------------------------------|
| YOLOv5   | You Only Look Once version 5                             |
| DCNFYOLO | Dual-Convolution Network and Feature Fusion based YOLOv5 |
| SAConv   | Switchable Atrous Convolution                            |
| DSConv   | Distribution Shifts Convolution                          |
| WIoU     | Wise-IoU                                                 |
| ECA      | Efficient Channel Attention                              |
| CBS      | ConvBNSiLU                                               |
| FPN      | Feature Pyramid Network                                  |
| SPPF     | Spatial Pyramid Pooling-Fast                             |

## References

- Huang, J.; Zhou, J.; Yang, H. A Small-Target Forest Fire Smoke Detection Model Based on Deformable Transformer for End-to-End Object Detection. *Forests* **2023**, *14*, 162. [\[CrossRef\]](#)
- Li, J.; Xu, R.; Liu, Y. An Improved Forest Fire and Smoke Detection Model Based on YOLOv5. *Forests* **2023**, *14*, 833. [\[CrossRef\]](#)
- Zhao, C.; Shu, X.; Yan, X.; Zuo, X.; Zhu, F. RDD-YOLO: A modified YOLO for detection of steel surface defects. *Measurement* **2023**, *214*, 112776. [\[CrossRef\]](#)
- Zhao, Z.; Zheng, P.; Xu, S.; Wu, X. Object detection with deep learning: A review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [\[CrossRef\]](#)
- Wu, X.; Sahoo, D.; Hoi, S.C. Recent advances in deep learning for object detection. *Neurocomputing* **2020**, *396*, 39–64. [\[CrossRef\]](#)
- Tao, H.; Lu, M.; Hu, Z. Attention-Aggregated Attribute-Aware Network With Redundancy Reduction Convolution for Video-Based Industrial Smoke Emission Recognition. *IEEE Trans. Ind. Inform.* **2022**, *7653*–7664. [\[CrossRef\]](#)
- Cao, Y.; Tang, Q.; Wu, X.; Lu, X. EFFNet: Enhanced Feature Foreground Network for Video Smoke Source Prediction and Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 1820–1833. [\[CrossRef\]](#)
- Yuan, F.; Zhang, L.; Xia, X.; Huang, Q.; Li, X. A Gated Recurrent Network With Dual Classification Assistance for Smoke Semantic Segmentation. *IEEE Trans. Image Process.* **2021**, *30*, 4409–4422. [\[CrossRef\]](#)
- Yin, H.; Chen, M.; Fan, W.; Jin, Y. Efficient Smoke Detection Based on YOLOv5s. *Mathematics* **2022**, *10*, 3493. [\[CrossRef\]](#)
- Huo, Y.; Zhang, Q.; Lin, G. A Deep Separable Convolutional Neural Network for Multiscale Image-Based Smoke Detection. *Fire Technol.* **2022**, *58*, 1445–1468. [\[CrossRef\]](#)
- Jing, T.; Zeng, M.; Meng, Q. SmokePose: End-to-End Smoke Keypoint Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 5778–5789. [\[CrossRef\]](#)
- Yin, Z.; Wan, B.; Yuan, F. A Deep Normalization and Convolutional Neural Network for Image Smoke Detection. *IEEE Access* **2017**, *5*, 18429–18438. [\[CrossRef\]](#)
- Yuan, F.; Zhang, L.; Xia, X. Deep smoke segmentation. *Neurocomputing* **2019**, *357*, 248–260. [\[CrossRef\]](#)
- Saponara, S.; Elhanashi, A.; Gagliardi, A. Real-time video fire/smoke detection based on CNN in antifire surveillance systems. *J. Real-Time Image Process.* **2021**, *18*, 889–900. [\[CrossRef\]](#)
- Tao, H.; Lu, X. Smoke Vehicle Detection Based on Spatiotemporal Bag-Of-Features and Professional Convolutional Neural Network. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 3301–3316. [\[CrossRef\]](#)
- Li, X.; Chen, Z.; Wu, Q.J.; Liu, C. 3D Parallel Fully Convolutional Networks for Real-Time Video Wildfire Smoke Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 89–103. [\[CrossRef\]](#)
- Verma, S.; Kaur, S.; Rawat, D.B. Intelligent Framework Using IoT-Based WSNs for Wildfire Detection. *IEEE Access* **2021**, *9*, 48185–48196. [\[CrossRef\]](#)
- Kaur, K.; Garg, S.; Kaddoum, G.; Ahmed, S.H. KEIDS: Kubernetes-Based Energy and Interference Driven Scheduler for Industrial IoT in Edge-Cloud Ecosystem. *IEEE Internet Things J.* **2020**, *7*, 4228–4237. [\[CrossRef\]](#)
- Zhang, J.; Xu, C.; Gao, Z. Industrial Pervasive Edge Computing-Based Intelligence IoT for Surveillance Saliency Detection. *IEEE Trans. Ind. Inform.* **2021**, *17*, 5012–5020. [\[CrossRef\]](#)

20. Gao, Z.; Xu, C.; Zhang, H. Trustful Internet of Surveillance Things Based on Deeply Represented Visual Co-Saliency Detection. *IEEE Internet Things J.* **2020**, *7*, 4092–4100. [\[CrossRef\]](#)
21. Połap, D.; Woźniak, M. Meta-heuristic as manager in federated learning approaches for image processing purposes. *Appl. Soft Comput.* **2021**, *113*, 107872. [\[CrossRef\]](#)
22. Gao, Z.; Zhang, H.; Dong, S. Salient Object Detection in the Distributed Cloud-Edge Intelligent Network. *IEEE Netw.* **2020**, *34*, 216–224. [\[CrossRef\]](#)
23. Pawar, A. A multi-disciplinary vision-based fire and smoke detection system. In Proceedings of the IEEE 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 5–7 November 2020; pp. 900–904. [\[CrossRef\]](#)
24. Khan, S.; Muhammad, K.; Hussain, T.; Ser, J.D. DeepSmoke: Deep learning model for smoke detection and segmentation in outdoor environments. *Expert Syst. Appl.* **2021**, *182*, 115125. [\[CrossRef\]](#)
25. Khan, M.; Khan, S.; Palade, V. Edge Intelligence-Assisted Smoke Detection in Foggy Surveillance Environments. *IEEE Trans. Ind. Inform.* **2020**, *16*, 1067–1075. [\[CrossRef\]](#)
26. Almeida, J.S.; Huang, C.; Nogueira, F.G. EdgeFireSmoke: A Novel Lightweight CNN Model for Real-Time Video Fire–Smoke Detection. *IEEE Trans. Ind. Inform.* **2022**, *18*, 7889–7898. [\[CrossRef\]](#)
27. Khan, S.; Khan, M.; Mumtaz, S. Energy-Efficient Deep CNN for Smoke Detection in Foggy IoT Environment. *IEEE Internet Things J.* **2019**, *6*, 9237–9245. [\[CrossRef\]](#)
28. Feng, X.; Cheng, P.; Chen, F.; Huang, Y. Full-Scale Fire Smoke Root Detection Based on Connected Particles. *Sensors* **2022**, *22*, 6748. [\[CrossRef\]](#)
29. Mozaffari, M.; Li, Y.; Ko, Y. Real-time detection and forecast of flashovers by the visual room fire features using deep convolutional neural networks. *J. Build. Eng.* **2023**, *64*, 105674. [\[CrossRef\]](#)
30. Li, X.; Zhang, G.; Sanqing Tanand Yang, Z. Forest Fire Smoke Detection Research Based on the Random Forest Algorithm and Sub-Pixel Mapping Method. *Forests* **2023**, *14*, 485. [\[CrossRef\]](#)
31. Polap, D.; Jaszcz, A. Sonar Digital Twin Layer via Multiattention Networks With Feature Transfer. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–10. [\[CrossRef\]](#)
32. Du, X.; Cheng, H.; Ma, H.; Lu, W. DSW-YOLO: A detection method for ground-planted strawberry fruits under different occlusion levels. *Comput. Electron. Agric.* **2023**, *214*, 108304. [\[CrossRef\]](#)
33. Wu, C.; Lei, J.; Liu, W.; Ren, M. Unmanned Ship Identification Based on Improved YOLOv8s Algorithm. *Comput. Mater. Contin.* **2024**, *78*, 3071–3088. [\[CrossRef\]](#)
34. Hu, Y.; Zhan, J.; Zhou, G.; Chen, A. Fast forest fire smoke detection using MVMNet. *Knowl.-Based Syst.* **2022**, *241*, 108219. [\[CrossRef\]](#)
35. Wang, H.; Xu, Y.; He, Y. YOLOv5-Fog: A Multiobjective Visual Detection Algorithm for Fog Driving Scenes Based on Improved YOLOv5. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–12. [\[CrossRef\]](#)
36. Tu, X.; Yuan, Z.; Liu, B.; Liu, J. An improved YOLOv5 for object detection in visible and thermal infrared images based on contrastive learning. *Front. Phys.* **2023**, *11*, 1193245. [\[CrossRef\]](#)
37. Yang, J.; Zhu, W.; Sun, T.; Ren, X. Lightweight forest smoke and fire detection algorithm based on improved YOLOv5. *PLoS ONE* **2023**, *18*, e0291359. [\[CrossRef\]](#)
38. Al-Smadi, Y.; Alauthman, M.; Al-Qerem, A. Early Wildfire Smoke Detection Using Different YOLO Models. *Machines* **2023**, *11*, 246. [\[CrossRef\]](#)
39. Li, C.; Zhao, G.; Gu, D.; Wang, Z. Improved Lightweight YOLOv5 Using Attention Mechanism for Satellite Components Recognition. *IEEE Sens. J.* **2023**, *23*, 514–526. [\[CrossRef\]](#)
40. Song, Y.; Xie, Z.; Wang, X.; Zou, Y. MS-YOLO: Object Detection Based on YOLOv5 Optimized Fusion Millimeter-Wave Radar and Machine Vision. *IEEE Sens. J.* **2022**, *22*, 15435–15447. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.