IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 March 2023, accepted 26 March 2023, date of publication 31 March 2023, date of current version 7 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3263479

APPLIED RESEARCH

Flame and Smoke Detection Algorithm Based on ODConvBS-YOLOv5s

JINGRUN MA, ZHENGWEI ZHANG, WEIEN XIAO, XINLEI ZHANG, AND SHAOZHANG XIAO
Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huai'an, Jiangsu 223003, China

Corresponding author: Zhengwei Zhang (zhangzhengwei@hyit.edu.cn)

This work was supported in part by the National Statistical Science Research Project under Grant 2018LY12, and in part by the Opening Project of Guangdong Province Key Laboratory of Information Security Technology under Grant 2020B1212060078.

ABSTRACT Real-time and accurate detection of flame and smoke is an important prerequisite to reduce the loss caused by fire. There are exists some problems in traditional detection algorithms of flame and smoke, such as low accuracy, high miss rate, low detection efficiency, and low detection rate of small targets. This paper proposes the detection algorithm of flame and smoke based on ODConvBS in YOLOv5s. Firstly, the ordinary convolutional blocks in the backbone network of YOLOv5s are replaced with ODConvBS to achieve the extraction of attentional features from the convolutional kernel. Secondly, Gnconv is introduced into Neck to improve the high-order spatial information extraction ability of the model. Then, the Shuffle Attention module is added at the end of the Neck to facilitate the fusion of different groups of features. Finally, the prediction part uses a SIOU loss function that can take into account the angle of the prediction frame vectors to accelerate the model convergence. When utilizing the self-made dataset of flame and smoke, the upgraded YOLOv5s model mAP grew by 9.3%. At the same time, the accuracy rate, the recall rate, and the detection speed increased to 83.5%, 83.7%, 33.3FPS respectively.

INDEX TERMS YOLOv5s, object detection, Gnconv, attention mechanism, ODConvBS.

I. INTRODUCTION

Fire will inflict significant damage to human life and property in daily activities, as well as do harm to the healthy development of society. However, in the early stages of a fire disaster, the flame is easily extinguished. As a result, by detecting flame and smoke accurately and quickly, the loss caused by the fire may be minimized to sustain normal production. Early flame detection frequently collects flame and smoke data using different temperature sensors, smoke sensors, and photosensitive sensors to assess whether a fire has occurred. However, the installation position and effective range of the sensor, as well as the external light and ambient humidity, will have a significant impact on the detection accuracy of the flame and smoke.

The task of object detection is to find and classify all target objects in a picture, which is one of the fundamental tasks in the computer vision field. At the current stage, object detection algorithms are divided into two categories: Twostage [1] and Onestage [2]. The twostage architecture generates pre-selected boxes that may contain objects to be detected and extracted by features and then conducts classification and regression localization. The twostage architecture does not need to generate pre-selected boxes and can extract features directly in the network to predict the classification and location of an object

This paper has four main contributions:

1: The ODConvBS module is suggested and implemented in the algorithmic backbone network to minimize computational complexity and improve the ability to extract features of the multi-convolutional kernel fusion model.

2: The Gnconv-FPN pyramid structure, which is based on recursive gated convolution, can increase the model's high-order spatial interaction capabilities, minimize missed detection rates, and improve the detection accuracy of small objects.

3: The Shuffle Attention module is introduced to quickly locate the region of interest and suppress the useless information extracted from the network.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiju Poovvancheri.

4: Using the SIOU loss function to account for the vector angle between regressions and adding matching directions to the original to speed up model convergence.

II. RELATED WORK

With the continuous upgrading of computer vision algorithms and hardware conditions, deep learning-based methods for detecting flame and smoke have surpassed traditional manual methods, and deep learning models can extract more abstract and deeper features from images with more powerful generalization compared to traditional methods. In 2010, Frizzi et al [3] first used a convolutional neural network for the image of flame and smoke detection, which pioneered the feature extraction algorithm of flame and smoke The deep learning-based flame and smoke detection task can be divided into three parts: classification [4] (determining whether the input image contains flame or smoke), detection (identifying whether the image contains flame and smoke and annotating it with an anchor), and segmentation [5] (identifying whether the image contains flame and smoke and annotating its shape). In 2019, Lin et al [6] developed a joint detection framework by combining Faster R-CNN and 3D CNN, where Faster R-CNN enables smoke localization in static spatial information and 3D CNN [7] achieves the recognition of smoke by combining the information of dynamic spatiotemporal. Compared with the common convolutional detection algorithm of smoke, this method improves significantly in the detection accuracy of smoke. In 2020, Li et al [8] used Faster R-CNN [9], R-FCN [10], SSD [11], and YOLOv3 [12] for flame detection, and they found that the CNN-based flame detection model could achieve a better balance of accuracy and detection. In 2021, Saponara et al. [13] deployed the YOLOv2 lightweight neural network to an embedded mobile device so that real-time detection of flame and smoke on the spot could be achieved. In 2021, Torabian et al [14] used a fire detection algorithm using motion analysis with fractal and spatiotemporal features to provide RGB probabilistic models to separate moving regions that have similar colors to the fire regions in each frame. Spatiotemporal features such as correlation coefficients and mutual information are then extracted from the candidate regions. In 2022, Avazov et al [15] ran the YOLOv4 algorithm on a three-layer Banana Pi M3 board that could provide an alarm within 8 seconds of a fire outbreak. In 2022, Majid et al [16] trained images of flame to detect flames with the help of a transfer learning strategy to visualize and localize flame in images using the Grad-CAM method. In 2022, Xue et al [17] adapted SPPF to SPPFP in the YOLOv5s backbone, allowing the model to better extract global information about small targets of flame. In 2022, Hu et al [18] proposed a value-transformed attention mechanism that utilizes the color and texture feature information of smoky images to further enhance the weight distribution of texture information. Also proposed Mixed-NMS which can consider angular information and centroid distance. In 2022, Li et al [19] designed a DLFR Module to reduce information loss during the acquisition of smoke images and proposed a hybrid attention module to reduce the interference of noisy images on model accuracy. In 2022, Khudayberdiev et al [20] proposed Light-FireNet, a lightweight deep learning framework inspired by the combination of lightweight convolution mechanisms in H-Swish. In 2022, Hosseini et al [21] proposed UFS-Net capable of identifying fire hazards by classifying video frames into eight categories, improving the reliability of the model by using a decision module based on the voting scheme. Object detection has also been widely employed in the field of autonomous driving in recent years, which is important for automobiles' automated recognition of barriers and traffic signs. In 2022, Liang et al. [22] and colleagues suggested an enhanced sparse R-CNN method for detecting traffic signs during unmanned driving. Our upgraded system can recognize traffic signs more quickly and precisely. In 2022, Gu et al. [23] and colleagues suggested a novel lightweight framework based on YOLOv4 to recognize traffic indicators. This new framework successfully decreases the algorithm's computational cost while also improving the model's generalization and resilience. In 2022, Wang et al. [24] and colleagues deployed the upgraded YOLOv5 deep learning network for real-time multi-scale traffic sign identification. The revised deep learning algorithm has better universality and superiority after experiments.

However, there are still some with the existing algorithms such as low detection accuracy, slow detection speed, high leakage rate, and slow convergence speed. Therefore, an improved YOLOv5s flame smoke target detection algorithm based on ODConvBS is proposed and improved according to the existing problems. Firstly, in order to improve the generalization of the model, the Mosaic [25] mosaic data enhancement technique and Mixup [26] mixed class data enhancement technique are used to further enrich the data diversity and improve the robustness of the model. Secondly, in a bit to improve the model's speed and accuracy for flame and smoke detection, Omni-dimensional dynamic convolution (ODConv) is added to the backbone network's convolutional block to form a new convolutional block (ODConvBS), which reduces network computation and improves the multi-convolutional kernel fusion model expression capability. To address the model's high miss detection rate, recursive gated convolution (Gnconv) is introduced into FPN to form a new Gnconv-FPN structure, which improves the model's interaction ability of higher-order information, achieves the same effect as the self-attention mechanism, avoids target information loss, and improves the detection accuracy of small targets even further. To boost the ability to extract features of the model even further, an ultra-lightweight replacement attention mechanism (Shuffle Attention) is included after the FPN structure to integrate all features and perform component feature communication through channel replacement operation. Lastly, the SIOU [27] loss function is used to enhance the training and convergence time of the model by fully accounting for the vector angle between regressions.

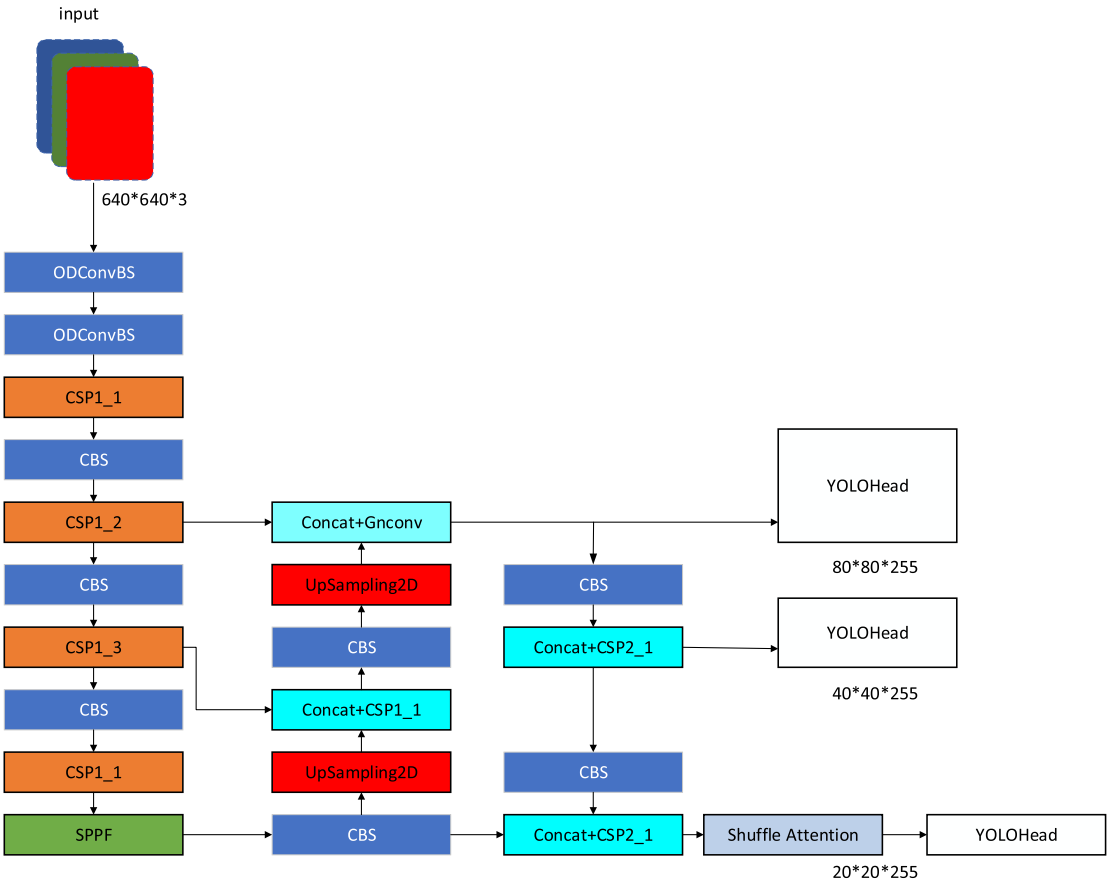


FIGURE 1. Improved YOLOv5s network.

III. THE BASIC STRUCTURE OF YOLOv5s

YOLOv5s has been upgraded to version 6.2. The backbone component contains CBS down-sampling processing module, CSP1 structure, and SPPF (spatial pyramid pooling) The neck uses FPN [28] (characteristic pyramid) network. The prediction of part uses CIOU [29] as the loss function by

$$y = (\alpha_{w1} \odot \alpha_{f1} \odot \alpha_{c1} \odot \alpha_{s1} \odot W_1 + \ldots + \alpha_{wn} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) * x \quad (1)$$

default and offers three distinct output scales.

The reasoning speed of the YOLOv5s model is extremely excellent for the application scenario of flame and smoke detection. However, the accuracy of flame and smoke detection has to be refined and enhanced. The following is algorithm flaws:

1. The multiple CONV and CSP modules are employed in the backbone network and feature pyramid structure, which quickly leads to feature map redundancy and impacts detection accuracy and speed.

2. The target missed detection rate is considered when employing YOLOv5s for target recognition on flame and smoke images.

IV. PROPOSED METHOD

A complete flame and smoke detection model is built on the computer as shown in Figure 1. The input flame and smoke images first go through feature extraction based on the ODConvBS backbone network. At the end of the backbone network, the SPPF module, which is faster, is used to unify the scale of the feature maps extracted by the backbone network and improve the accuracy of the feature extraction. The feature maps are then sent to the neck network (Gnconv-FPN) for feature processing and fusion, enabling the interaction of high-order spatial information in the feature maps and achieving the effect of self-attention feature extraction. At the end of the neck network, the SA module is used to promote information fusion between different groups. Finally, the information is sent to the head network to complete the object detection. The improved network model structure is shown in Figure 1.

A. YOLOv5s BACKBONE NETWORK BASED ON ODConvBS

By the use of attention mechanisms in the convolutional kernel, dynamic convolution can improve the performance of CNN networks. ODConv [30] (Omni-dimensional dynamic convolution) uses dynamic convolution to create a more diversified and effective attention mechanism, which is then

inserted into the convolutional kernel space. It employs a unique attention approach to learn convolutional kernel features in parallel throughout all four dimensions of the convolutional kernel space. These four types of attention mechanisms complement each other, and applying these four attention mechanisms to convolution kernels can further enhance the feature extraction capabilities of CNN. When compared with other dynamic convolution algorithms, ODConv has only one convolutional kernel and a far lower number of parameters. ODConv can ensure efficiency while maintaining accuracy, and its generalization ability is strong enough to fulfill the demands of flame and smoke detection.

The ODConv calculation along the convolution kernel. Equation 1 shows how the formula is defined, contains four dimensions: the position multiplication operation in the space dimension the channel multiplication operation in the input channel dimension, the filter multiplication operation in the output channel dimension, and the kernel multiplication operation in the convolution kernel space

Where α_{wi} represents the attention scalar for the entire convolution kernel, α_{fi} represents the attention scalar for the output channel, α_{ci} represents the attention scalar for the input channel, α_{si} represents the attention scalar for the convolution kernel space, W_i represents a convolution kernel.

The ODConvBS module is divided into three parts. The first part is ODConv, which can conduct all-around feature extraction on the convolutional kernel space. The second part is Batch Normalization, which can avoid gradient expansion and disappearance, and the last part is SiLU activation function which equalizes large-valued gradients. Figure 2 depicts the ODConvBS structure.

FIGURE 2. ODConvBS structure diagram.

The first two CBS modules in the YOLOv5s backbone serve as input pictures for size reduction and channel tuning. By replacing these two modules with ODConvBS, picture information can be retrieved from the backbone network of the convolutional kernel using the attention technique while taking the original function into account, improving the characteristic aggregation capacity of the backbone network. Figure 3 depicts the enhanced backbone of YOLOv5s.

B. SHUFFLE ATTENTION (SA) MECHANISM THAT EFFICIENTLY COMBINES SPATIAL AND CHANNEL INFORMATION

Group convolution, a spatial attention mechanism, a channel attention mechanism, and ShuffleNetV2 [32] are all combined in the design concept of SA [31]. The Tensor is first divided into g groups, and then each group is internally processed using a SA Unit. The diamond-shaped square in Figure 4 is implemented similarly to SE, while the circular

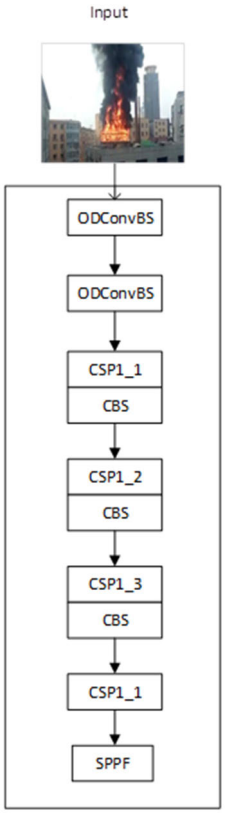


FIGURE 3. YOLOv5s backbone network structure diagram based on ODConvBS.

square in Figure 4 represents the spatial attention mechanism in the SA, which is done by the use of GN. The SA Unit incorporates Concatenation to merge information within a group and then employs Channel Shuffle to reorganize the group, allowing information to flow between different entities.

In the upgraded version of YOLOv5s, the SA module is positioned at the end of the FPN architecture. By performing feature extraction and combining different sets of information via the FPN, the network can obtain more diverse and comprehensive information, which facilitates the subsequent object detection task. The structure of the SA network is illustrated in Figure 4.

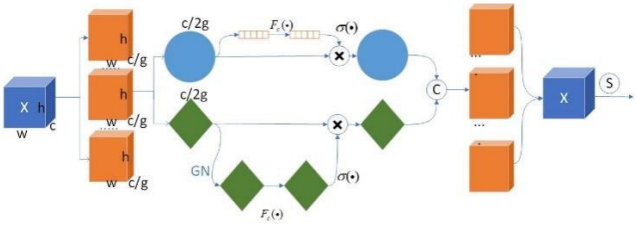


FIGURE 4. SA attention structure diagram.

C. SPATIAL PYRAMID POOLING MODULE SPPF

Fixed-input design is required in convolutional neural networks, and spatial pyramidal pooling can help us achieve this.

SPPF is a fast variant of SPP that is twice as quick while delivering the same computational results.

In the SPPF structure, the feature map first goes through CBS (Conv + BN + SiLU) and then enters three maximum pooling layers of 5 × 5 size in turn, and then the results of these three maximum pooling layers are summed up, and finally, the feature vector map extracted from the backbone network is scaled uniformly by the CBS module at the end of the network structure to ensure that the object location and size of the feature map. Figure 5 depicts the SPPF structure.

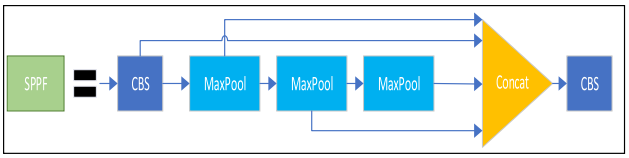


FIGURE 5. SPPF structure.

D. IMPROVED FEATURE MAP PYRAMID NETWORK GNCONV-FPN

The FPN of Pyramid networks with feature map primarily address the multiscale problem in object identification and can enhance the detection performance of small targets by modifying network connections with almost no increased computational cost in the original model. However, FPN cannot realize adaptive input, large-scale and high-order spatial information interaction, while Gnconv [33] can make up for these shortcomings. Gnconv is a convolution operation that can achieve large-range and high-order spatial interactions. It is built with standard convolution, linear projection, and element multiplication, but it has an input adaptive space similarity to the transformer [34] blending function to achieve the effect of self-attentive feature extraction. Figure 6 depicts the Gnconv structure.

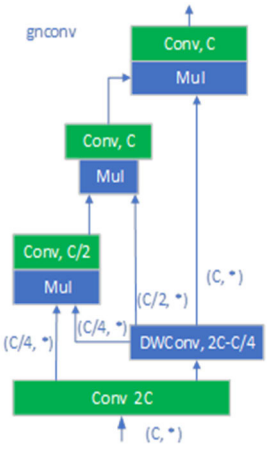


FIGURE 6. Gnconv structure diagram.

Gnconv-FPN replaces the CBS in the Neck part of YOLOv5s with Gnconv recursive gate convolution. It implements high-order spatial information interaction on the feature map before fusing with the connection group information, further expanding the receptive field, which is conducive to further feature extraction in the Neck part and subsequent prediction.

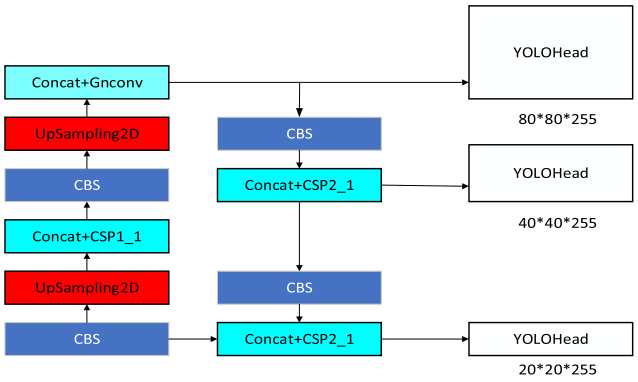


FIGURE 7. Gnconv-FPN structure diagram.

E. SIOU LOSS FUNCTION

The weighted sum of Classification_Loss, Localization_Loss, and Confidence_Loss forms the total loss of YOLOv5s network, and the attention of each network to different losses can be adjusted by changing the weights In the realm of object detection Boundary box prediction plays an important role in the field of object detection. If you want to frame the target in the object detection task, you need to predict the location data of the boundary box. The squared loss is used in the first variant, as illustrated in Equation 2.

$$L_{local} = (x - x^*)^2 + (y - y^*)^2 + (w - w^*)^2 + (h - h^*)^2 \tag{2}$$

x^*, y^*, w^*, h^* are the coordinates and width and height of the upper-left corner of the real box, and x, y, w, z are the coordinates and width and height of the predicted box, respectively. Boundary box prediction needs to focus on the overlap area between the predicted box and the real box, and a higher ratio of overlap area to the union area of the two boxes indicates a better prediction. However, using squared difference loss cannot effectively measure this

The CIOU loss calculation used by YOLOv5s has several flaws as well: the aspect ratio indicates the relative value, and there is some uncertainty; the balance of tough and simple samples is not taken into account. SIOU appeared at a historic period to answer the difficulty of IOU itself.

The SIOU loss function takes the vector angle between the needed regressions into account, then adds the matching direction on the original basis and redefines the model penalty index. This redesigned penalty index may significantly speed up the training convergence process and effect, causing the prediction box to swiftly shift to the nearest axis, and the subsequent approach just requires a regression of coordinate X or Y. Angle cost, Distance cost, Shape cost, and IOU cost comprise the SIOU loss function.

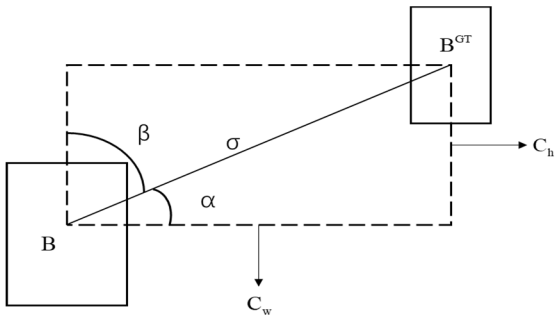


FIGURE 8. Schematic diagram of angle contribution of SIOU loss function.

Angle cost can minimize the number of distance-related variables by adding this angle-aware LF component. Basically, the model will try to bring the prediction to the X or Y axis first (whichever is closest), and then continue to approach along the relevant axis. Figure 8 depicts the method for determining the angle contribution to the loss function.

If $\alpha \leq \pi/4$, the convergence process will first minimize α, otherwise minimize β: $\beta = \pi/2 - \alpha$. To achieve the previous point, the loss function components are defined using Equation 3 below.

$$\begin{aligned} \Lambda &= 1 - 2 \times \sin^2(\arcsin(x) - \frac{\pi}{4}) \\ x &= \frac{c_h}{\sigma} = \sin(\alpha), \quad \sigma = \sqrt{(b^{gt}_{c_x} - b_{c_x})^2 + (b^{gt}_{c_y} - b_{c_y})^2}, \\ c_h &= \max(b^{gt}_{c_y}, b_{c_y}) - \min(b^{gt}_{c_y}, b_{c_y}). \end{aligned} \tag{3}$$

Distance cost is redefined based on Angle cost. Such as formula 4.

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) \tag{4}$$

In $\rho_x = (\frac{b^{gt}_{c_x} - b_{c_x}}{c_w})^2$, $\rho_y = (\frac{b^{gt}_{c_y} - b_{c_y}}{c_h})^2$, $\gamma = 2 - \Lambda$.

Equation 3 shows that when $\alpha \to 0$ the contribution of Distance cost is considerably decreased. On the contrary, the bigger the share of Distance cost, the closer it is to $\pi/4$. As the angle rises, the challenge becomes more difficult. As the angle grows, γ is allocated a time-priority distance value.

Shape cost is defined in Equation 5.

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^{\theta} \tag{5}$$

In $\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}$, $\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$.

The value of θ specifies the Shape cost, and it is unique for each dataset. The value of θ is a critical component in this equation because it determines how much attention is paid to Shape cost. If the value of θ is set to 1, it will quickly optimize a Shape, sacrificing the Shape's free mobility. To determine the value of θ, the author uses a genetic algorithm for each data set, and the value of θ is empirically near 4, and the author sets a range of 2 to 6 for this parameter in the work.

The IOU calculation is shown in Figure 9.

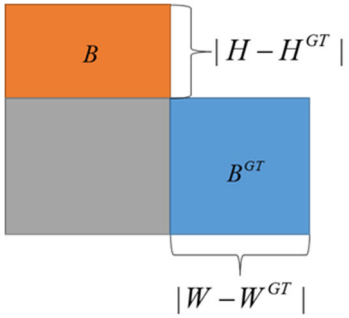


FIGURE 9. IOU calculation.

Finally, the loss function is defined as shown in Equation 6.

$$L_{box} = 1 - IOU + \frac{\Delta + \Omega}{2} \tag{6}$$

In $IOU(B, B_{gt}) = \left|\frac{B \cup B_{gt}}{B \cap B_{gt}}\right|$.

The SIOU final loss is defined as shown in Equation 7.

$$L = W_{box} L_{box} + W_{cls} L_{cls} \tag{7}$$

where L_{cls} is the focused loss, W_{box} and W_{cls} are the frame and classification loss weights. A genetic method is employed to compute W_{box}, W_{cls}, and θ. To train the genetic algorithm, a small subset of the training set is selected and the values are computed until a value less than a threshold is reached or the maximum number of iterations permitted is reached.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL ENVIRONMENT

The network model was created with PyTorch version 1.9, Python version 3.8, an Ubuntu system, and a Tesla V100-SXM2 graphics card with 16G RAM. The pre-training weights are the official YOLOv5s weights, the training generation is 100 epochs, the Batch size is 16, the initial learning rate is 0.01, and the starting momentum of SGD is 0.097.

B. DATA ENHANCEMENT AND DIVISION

Due to the small number of images, single scene and low resolution in the public of flame and smoke dataset, it is not conducive to improving the generalization ability of the model. To reflect the improved model generalization and small target detection ability, it is necessary to further improve the diversity of datasets. Therefore, this paper crawls the flame and smoke images on the network through crawlers and then labels the images into data sets to train and evaluate the model. The dataset of 4998 photos was separated into training, validation, and test sets in the ratio of 8:1:1, covering a range of flame smoke scenes, following the study topic of this work.

C. DATA ANALYSIS

Indicators such as P, MAP, and FPS are required for assessment based on the YOLOv5s model's improved impact.

TABLE 1. Ablation experiments.

Number	ODConv	SA	Gnconv_FPN	SIOU	Precision	Recall	mAP	mAP.5:0.95
YOLOv5s					75.8	74.4	78.3	45.5
2	√				80.2	75.3	81	51.4
3		√			77.7	86.2	84.9	54
4			√		78.9	77.7	81.6	51.5
5				√	77.2	85.9	85.1	54
6	√			√	79.6	84.4	85.5	54.4
7		√		√	80.4	81.7	84.8	54
8			√	√	78.1	87.3	85	53.2
ours	√	√	√	√	**83.5**	**83.7**	**87.6**	**57.9**

Because this is a dual-objective experiment and the average precision is denoted by mAP.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$AP = \int_0^1 Precision(t)dt \tag{10}$$

$$mAP = \frac{1}{N}\sum AP_i \tag{11}$$

where TP is the number of correctly recognized positive samples, FP is the number of correctly detected negative samples, and FN is the number of backgrounds wrongly detected as positive samples. The number of frames per second (FPS) conveyed shows the number of pictures that the algorithm can process per second. ODConv is an abbreviation for Omni-dimensional dynamic convolution. SA is an abbreviation for ShuffleAttention attention mechanism and Gnconv is an abbreviation for recursive gated convolution, and SIOU is an abbreviation for loss function. Table 1 depicts the ablation trials.

As shown in Table 1, adding ODConv to the backbone of YOLOv5s can increase the map by 2.7% Adding the Shuffle Attention mechanism at the end of the neck can increase the mAP by 6.6% Adding the Gnconv FPN pyramid structure can increase the mAP by 3.3% Adding the SIOU loss function can increase the mAP by 3.3%. The map is then enhanced by 6.8%. When all of the improvement strategies are applied to the YOLOv5s model at the same time, the mAP increases by 9.3% over the original model, the mAP0.5:0.95 increases by 12.4%, the accuracy rate increases by 7.7%, and the recall rate increases by 9.3%, demonstrating the network's superiority.

On the self-created flame and smoke dataset, the enhanced flame and smoke detection model is compared to other popular object detection methods in this research. Table 2 displays the experimental outcomes.

TABLE 2. Comparison results of mAP of different algorithms on the self-made flame and smoke dataset.

Model	Fire-AP	Smoke-AP	mAP
SSD	58.78	47.87	55.3
Faster R-CNN	65.39	46.23	55.8
YOLOV3	69.8	38.1	50
YOLOV4	74.5	52.4	63.4
YOLOv5s	84.7	71.8	78.3
YOLOv5x6+ TTA	89.9	74.6	82.3
Our	93.5	81.6	87.6

According to Tables 2 and 3, this article suggests that YOLOv5s based on ODConvBS outperform the two-stage target identification technique Faster R-CNN in terms of mAP. The mAP has risen by 37.6%, 24.2%, 9.3%, and 5.3%, respectively, when compared to the monocular target detection algorithms YOLOv3, YOLOv4, YOLOv5s, and YOLOv $5 \times 6 +$ TTA.

Because it is difficult for the single-layer feature map produced by Faster RCNN to tackle multi-scale issues, the model of accuracy for flame and smoke detection would suffer. When the 50% region of the object picture is utilized as the recognized standard, YOLOv3 has the best accuracy. However, when the standard rises, so does the accuracy. Both YOLOv4 and YOLOv5 employ the CIOU loss function, which has an uncertain aspect ratio and does not account for the balance of tough and easy samples, resulting in slower convergence. MobileOne introduces a simple reparameterization branch in the training phase which effectively reduces the number of model parameters, and the model detection accuracy is better than the YOLOv5 model, but the detection accuracy for flame and smoke is lower than the improved algorithm proposed in this paper.

The GnconvFPN structure is used in the improved method suggested in this research, which can conduct high-order spatial interaction under the premise of solving the multi-scale problem in object identification, producing a comparable impact to self-attention without generating extra calculations.

TABLE 3. Comparison results of different algorithms in accur-, acy speed, and computational complexity on the self-made flame and smoke dataset.

Model	mAP	Parameters	GFLOPs	FPS
YOLOV3	50	62.55M	155.6	9.9
YOLOV4	63.4	9.1M	20.6	23.81
YOLOv5s	78.3	7.01M	15.9	31.25
YOLOv5x6+TTA	82.3	86M	203.8	8.26
MobileOne+FPN	79.4	6.85M	16.9	28.46
Our	87.6	7.2M	14.8	33.3

By utilizing the ODConvBS module in the backbone network of the original model, the dynamics in dimensions such as airspace, input channels, and output channels may be evaluated concurrently, improving the accuracy of the model. The SA attention mechanism is employed near the conclusion of Neck to merge spatial and channel attention and make the model more efficient while acquiring picture input. Finally, the SIOU loss function is employed to accelerate model convergence. Furthermore, as shown in Table 3, the flame and smoke algorithm presented in this study has a fast speed of detection while maintaining accuracy. Figure 10 can more intuitively see the performance comparison between the improved model and other mainstream models. When the abscissa is mAP, the higher the value, the higher the detection accuracy. When the abscissa is FLOPs, the lower the value, the lower the amount of floating-point calculations. When the abscissa is FPS, the higher the value, the faster the detection speed. When the abscissa is Params, the lower the value, the lower the number of model parameters. It can be seen from Figure 10 that the accuracy of the improved model, the amount of floating-point calculations, and the detection speed are all better than other models.

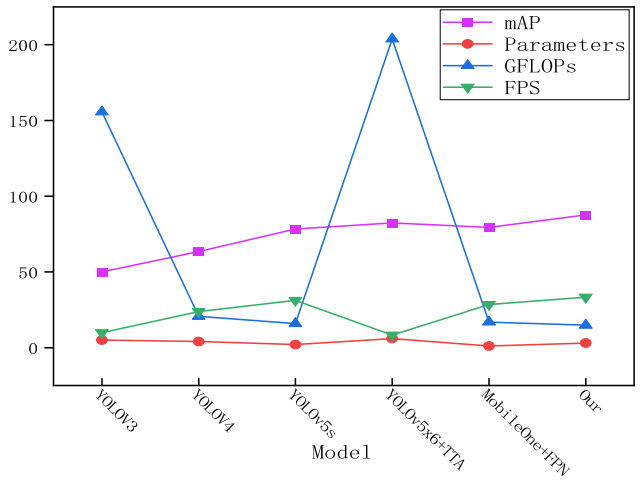


FIGURE 10. Loss curve of each model.

The model improvement effect is shown in Figure 11. The detection effect of the original YOLOv5 is shown on the left, and the detection effect after improvement is shown on the right.

FIGURE 11. Comparison chart of experimental results (The image on the left is before the algorithm improvement, and the image on the right is after the algorithm improvement).

Compared with the pre-improved flame and smoke recognition accuracy, the improved model has been greatly improved. After the ODConvBS module extracts the attention feature of the convolution kernel, the deep learning model can better notice the target in the picture and better extract the flame and smoke features. Figure 11 shows that the original model has issues with missing detection and lower detection accuracy of small targets for flame and smoke detection, but after modification, the model's missing detection rate has fallen dramatically and its detection accuracy of small targets has been greatly improved. This is attributed to the Gnconv-FPN structure's capacity to improve the model's extraction of high-level feature semantic information from the input picture, completely capture the target in the image, and improve the detection performance of the model for small targets. Figure 11 also shows that the detection accuracy of the YOLOv5s model for small targets is significantly lower than that of the upgraded model. The prediction section utilizes SIOU as the loss function and employs three distinct output scales, which correspond to three different target predictions. This approach enhances the detection of small targets within the input image, resulting in more effective performance.

VI. SUMMARIZE

The use of classical image enhancement methods and deep learning algorithms such as YOLOv3 and v4 to recognize flames and smoke in fire scenes is difficult to meet practical application requirements. Therefore, this article proposes the use of YOLOv5s based on the ODConvBS model to recognize flames and smoke, further improving detection efficiency and accuracy. Therefore, the work suggests using an ODConvBS-based YOLOv5s method to

recognize flame and smoke to increase detection efficiency and accuracy. The following are the main contributions of this paper: 1) Propose the ODConvBS module based on Omni-dimensional dynamic convolution 2) Use Gnconv-FPN to improve the feature pyramid structure of the original network 3) Add the SA attention mechanism to improve the model's ability to extract image information. 4) Improve the prototype loss calculation by using the SIOU loss function.

The testing findings reveal that: 1) In the case of the best settings, the deep learning algorithm of upgraded YOLOv5s has a map of 87.6%, which is much better than the original YOLOv5s, SSD, Faster R-CNN, and other algorithms. 2) When compared to the original model, the updated model may consider detection efficiency while enhancing accuracy.

However, the algorithm presented in this article also has certain limitations: compared to the original model, it requires the extraction of high-order spatial information, resulting in an increased number of parameters and only a slight improvement in detection speed. In the future, we will explore lightweight backbone networks and new attention mechanisms to simplify the network, further improve detection efficiency and accuracy, and achieve a flame and smoke detection algorithm that balances both high precision and speed, thereby enabling real-time application in industrial scenarios.

REFERENCES

[1] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.

[2] X. Wu, D. Sahoo, and S. C. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, Jul. 2020.

[3] J. Chen, Y. He, and J. Wang, "Multi-feature fusion based fast video flame detection," *Building Environ.*, vol. 45, no. 5, pp. 1113–1122, May 2010.

[4] A. Pawar, "A multi-disciplinary vision-based fire and smoke detection system," in *Proc. 4th Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Nov. 2020, pp. 900–904.

[5] S. Khan, K. Muhammad, and T. Hussain, "Deepsmoke: Deep learning model for smoke detection and segmentation in outdoor environments," *Expert Syst. Appl.*, vol. 182, Nov. 2021, Art. no. 115125.

[6] G. Lin, Y. Zhang, G. Xu, and Q. Zhang, "Smoke detection on video sequences using 3D convolutional neural networks," *Fire Technol.*, vol. 55, no. 5, pp. 1827–1847, Sep. 2019.

[7] L. Wang, W. Li, and W. Li, "Appearance-and-relation networks for video classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1430–1439.

[8] P. Li and W. Zhao, "Image fire detection algorithms based on convolutional neural networks," *Case Stud. Thermal Eng.*, vol. 19, Jun. 2020, Art. no. 100625.

[9] S. Ren, K. He, and R. Girshick, "Towards real-time object detection with region proposal networks," 2019, *arXiv:1506.01497*.

[10] J. Dai, Y. Li, and K. He, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–11.

[11] W. Liu, D. Anguelov, and D. Erhan, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.

[12] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[13] S. Saponara, A. Elhanashi, and A. Gagliardi, "Real-time video fire/smoke detection based on CNN in antifire surveillance systems," *J. Real-Time Image Process.*, vol. 18, no. 3, pp. 889–900, 2021.

[14] M. Torabian, H. Pourghassem, and H. Mahdavi-Nasab, "Fire detection based on fractal analysis and spatio-temporal features," *Fire Technol.*, vol. 57, pp. 2583–2614, May 2021.

[15] K. Avazov, M. Mukhiddinov, F. Makhmudov, and Y. I. Cho, "Fire detection method in smart city environments using a deep-learning-based approach," *Electronics*, vol. 11, p. 73, Dec. 2022.

[16] S. Majid, F. Alenezi, S. Masood, M. Ahmad, E. S. Gündüz, and K. Polat, "Attention based CNN model for fire detection and localization in real-world images," *Expert Syst. Appl.*, vol. 189, Mar. 2022, Art. no. 116114.

[17] Z. Xue, H. Lin, and F. Wang, "A small target forest fire detection model based on YOLOv5 improvement," *Forests*, vol. 13, no. 8, p. 1332, Aug. 2022.

[18] Y. Hu, J. Zhan, and G. Zhou, "Fast forest fire smoke detection using MVMNet," *Knowl.-Based Syst.*, vol. 241, Jan. 2022, Art. no. 108219.

[19] J. Li, G. Zhou, and A. Chen, "Adaptive linear feature-reuse network for rapid forest fire smoke detection model," *Ecolog. Informat.*, vol. 68, May 2022, Art. no. 101584.

[20] O. Khudayberdiev, J. Zhang, and S. M. Abdullahi, "Light-FireNet: An efficient lightweight network for fire detection in diverse environments," *Multimedia Tools Appl.*, vol. 81, pp. 24553–24572, Mar. 2022.

[21] A. Hosseini, M. Hashemzadeh, and N. Farajzadeh, "UFS-Net: A unified flame and smoke detection method for early detection of fire in video surveillance applications using CNNs," *J. Comput. Sci.*, vol. 61, May 2022, Art. no. 101638.

[22] T. Liang, H. Bao, and W. Pan, "Traffic sign detection via improved sparse R-CNN for autonomous vehicles," *J. Adv. Transp.*, vol. 2022, pp. 1–16, Mar. 2022.

[23] Y. Gu and B. Si, "A novel lightweight real-time traffic sign detection integration framework based on YOLOv4," *Entropy*, vol. 24, no. 4, p. 487, Mar. 2022.

[24] J. Wang, Y. Chen, and Z. Dong, "Improved YOLOv5 network for real-time multi-scale traffic sign detection," *Neural Comput. Appl.*, vol. 35, pp. 7853–7865, Dec. 2022.

[25] F. Dadboud, V. Patel, and V. Mehta, "Single-stage UAV detection and classification with YOLOV5: Mosaic data augmentation and PANet," in *Proc. 17th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2021, pp. 1–8.

[26] C. Si, Z. Zhang, and F. Qi, "Better robustness by more coverage: Adversarial and mixup data augmentation for robust finetuning," in *Proc. Findings Assoc. Comput. Linguistics, (ACL-IJCNLP)*, 2021, pp. 1569–1576.

[27] Z. Gevorgyan, "SIoU loss: More powerful learning for bounding box regression," 2022, *arXiv:2205.12740*.

[28] T. Y. Lin, P. Dollár, and R. Girshick, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2117–2125.

[29] Z. Zheng, P. Wang, and D. Ren, "Enhancing geometric factors in model learning and inference for object detection and instance segmentation," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8574–8586, Aug. 2022.

[30] C. Li, A. Zhou, and A. Yao, "Omni-dimensional dynamic convolution," 2022, *arXiv:2209.07947*.

[31] Q. L. Zhang and Y. B. Yang, "SA-Net: Shuffle attention for deep convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 2235–2239.

[32] N. Ma, X. Zhang, and H. T. Zheng, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 116–131.

[33] Y. Rao, W. Zhao, Y. Tang, J. Zhou, S.-N. Lim, and J. Lu, "HorNet: Efficient high-order spatial interactions with recursive gated convolutions," 2022, *arXiv:2207.14284*.

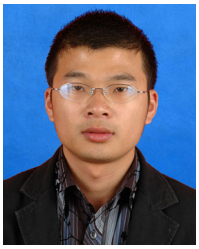
[34] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

JINGRUN MA received the B.S. degree from the Shandong Huayu Institute of Technology, in 2021. He is currently pursuing the degree with the Department of Computer and Software Engineering, Huaiyin Institute of Technology, Huai'an, China. His recent research interests include image processing and artificial intelligence.

XINLEI ZHANG received the B.S. degree in computer science and technology from the Jining Medical College, Rizhao, Shandong, China, in 2022. She is currently pursuing the degree with the Huaiyin Institute of Technology, Huai'an, Jiangsu, China. Her main research interest includes image segmentation.

ZHENGWEI ZHANG received the B.S. degree in computer science and technology from the Jiangsu University of Science and Technology, in 2004, the M.S. degree in computer science and technology from Jiangnan University, in 2011, and the Ph.D. degree in computer science and technology from the PLA University of Science and Technology, in 2017. His current research interests include information hiding and image processing.

WEIEN XIAO received the B.S. degree in computer science and technology from the Huaiyin Institute of Technology, in 2020, where he is currently pursuing the M.S. degree. His current research interests include information hiding, digital watermarking, and image processing.



SHAOZHANG XIAO received the B.S. degree in communication engineering and the M.S. degree in computer science and technology from Jiangnan University, in 2004 and 2013, respectively. His current research interests include information hiding and signal processing.