HighTech and Innovation Journal

ISSN: 2723-9535

Available online at www.HighTechJournal.org

HighTech and Innovation Journal

Vol. 5, No. 3, September, 2024

An Improved Fire Detection Algorithm Based on YOLOv8 Integrated with DGIConv, FourBranchAttention and GSIoU

Muxiang Zhang [1*]

[1] *School of Computer and Information Engineering, Henan University, Jinming Road, Kaifeng, Henan, 475004, China.*

Received 18 May 2024; Revised 21 August 2024; Accepted 26 August 2024; Published 01 September 2024

Abstract

Fire detection is highly important for people's lives and property, and enhancing its accuracy is essential. This study focused on utilizing and improving YOLOv8 to obtain higher detection accuracy for fire detection. Three methods were used. First, the newly designed DGIConv module replaces the original Conv module, thereby decreasing the computational complexity while enhancing the model's performance. Second, to enhance the recognition ability of flame targets, a new attention mechanism named FourBranchAttention was designed, and a comparison was made with other attention mechanisms. The experiments revealed that the newly designed attention mechanism performed best on the mAP50 and mAP50-95 metrics. Finally, to improve the convergence speed and localization ability of the model, the loss function is optimized by adopting better hyperparameters of the TaskAlignedAssigner and employing the newly designed GSIoU as an alternative to the original CIoU. Through ablation experiments, all three improvements improved the detection performance to a certain extent, and the model using the three improvements achieved the best performance. Compared with the baseline, the YOLOv8 model with DGIConv, FourBranchAttention, and the optimized loss function increased the mAP50 by 2.52% and the mAP50-95 by 3.37%. The mAP50 and mAP50-95 had reached 98.46% and 75.26%, respectively. Compared with previous models, such as SSD/YOLOv7, the performance metrics of enhanced YOLOv8 also exhibited significant enhancements, thereby augmenting the accuracy of fire detection.

Keywords: YOLOv8; Conv; Attention; Loss; IoU; mAP50.

1. Introduction

Fire has consistently proven to be a catastrophic force throughout history, possessing immense destructive capabilities and posing a significant threat to public safety. Hence, the implementation of fire detection systems is important. Conventional fire alarm systems often rely on infrared or smoke sensors; however, these sensors have inherent limitations in their detection ranges. In recent years, video cameras have been widely used, and fire detection technologies based on images have been continuously developed. This study investigated the use of image processing and deep learning in fire detection technologies.

The primary object detection research is conducted in two ways: two-stage algorithms, such as the RCNN series [1], and one-stage algorithms, such as the YOLO series [2, 3] and SSD [4]. The two-stage algorithm involves the generation of preselection boxes that potentially encompass objects and subsequently recognizes the targets within these preselection boxes. The one-stage technique directly predicts the location and categorization of objects. Han et al. [5] proposed a lightweight forest fire detection model named LUFFD-YOLO for UAV remote sensing data. The LUFFD-

* Corresponding author: yyk6588@qq.com

doi http://dx.doi.org/10.28991/HIJ-2024-05-03-09

YOLO model improves the existing YOLOv8n network by incorporating a set of optimizations. GhostNetv2 is utilized to enhance the conventional convolution of YOLOv8n. The ESDC2f architecture, which uses the MHSA attention mechanism, was proposed. The HFIC2f structure was redesigned by using the SegNeXt attention mechanism to enhance the integration of features from different layers. These improvements result in improved detection accuracy. The LUFFD-YOLO model yields an improvement in the mAP50 of 88.3%.

Cao et al. [6] proposed a novel detection technique based on an improved YOLOv5 model to enhance the visual representation of forest fires. It adds a plug-and-play global attention mechanism to improve the efficiency of neck and backbone feature extraction. Then, a reparameterized convolutional module is designed, and a decoupled detection head is used to accelerate the convergence speed. Finally, a weighted bidirectional feature pyramid network (BiFPN) is introduced to merge feature information. Through experiments on self-built forest and grassland datasets, the mAP50 reached 78.8%, and the mAP50-95 reached 49.0%.

Huang et al. [7] proposed an improved end-to-end deformable DETR model for forest fire smoke detection. The multiscale context contrasted local features, and a dense pyramid pooling module are used. Several dilated convolutions with different rates make full use of context information and local information of inconspicuous objects, which improves the performance of early forest fire smoke detection. It also proposes an iterative bounding box combination method to reduce the occurrences of false and missed detections and generate a bounding box for forest fire smoke more accurately to the ground truth. The model not only achieves high detection accuracy for smoke but can also detect early forest fire smoke, which is too small and inconspicuous to be detected by common models. The mAP50 reached 88.4%.

Wang et al. [8] proposed an EA-YOLO model in which an efficient attention mechanism is integrated into the backbone network, multichannel attention (MCA), and the number of parameters of the model is reduced by introducing the RepVB module. The study also designs a multiweighted multidirectional feature neck structure, the multidirectional feature pyramid network (MDFPN), to enhance the model's ability to fuse flame target feature information. The experimental results show that EA-YOLO's mAP50 achieves values of 84.9% and 81.5% on the Fire-Smoke and Ro-Fire-Smoke datasets, respectively, which are 6.5% and 7.3% higher than those of YOLOv7. Wei [9] proposed an object detection algorithm based on YOLOv8 with an improvement incorporating the SE attention mechanism. It achieves an average mAP50 value of 73% on the self-built dataset containing 2059 images.

Saydirasulovich et al. [10] presented an enhanced YOLOv8 model customized to the context of unmanned aerial vehicle (UAV) images and achieved heightened precision in terms of detection accuracy. The research incorporates Wise-IoU (WIoU) v3 as a regression loss for bounding boxes. The conventional convolutional process within the intermediate neck layer is substituted with the ghost shuffling convolution mechanism. This study introduces the BiFormer attention mechanism, which strategically directs the model's attention toward the feature intricacies of forest fire smoke, simultaneously suppressing the influence of irrelevant, nontarget background information. The obtained experimental findings highlight the enhanced YOLOv8 model's effectiveness in smoke detection, with the mAP50 of 79.4%, indicating a notable rise of 3.3%.

The above researches are based on various versions of YOLO and introduce different improvement methods. Through experiments, the model performance and detection accuracy have improved to a certain extent, and indicators such as mAP50 have been enhanced. However, these improvements are mostly modifications to network structures such as feature extraction modules and fusion modules or the use of existing attention mechanisms and loss functions, and the mAP50 metric rarely exceeds 90%. This study designs new loss functions and attention mechanisms to improve model performance and enhance important metrics such as mAP50/mAP50-95.

The YOLO model has undergone multiple updates and iterations by researchers since its introduction by Redmon et al. [11]. The 8th version of the YOLO algorithm, known as YOLOv8, was released in 2023, and YOLOv8 was adopted in this study to detect fires, which is enhanced to yield more favorable results. The YOLOv8 model comprises a comprehensive set of 21 units, including Conv and C2f. In general, the system can be categorized into two main components: a feature extraction backbone network and detection heads. Conv, C2F, and SPPF are critical components of the backbone network, and these modules serve as feature extraction. Conv adopts the Nwankpa et al. [12] activation function, C2f incorporates feature fusion and residual connections, and SPPF uses a spatial pyramid pooling structure, which remains the widely adopted PAN-FPN [13]. Downsampling is executed before the upsampling process, and the integration of upsampling and downsampling through cross-layer fusion leads to three distinct detection heads.

The detection head adopts a decoupled structure to effectively segregate classification and detection tasks, and simultaneously, the anchor-based approach translates to an anchor-free approach [14, 15] in which the anchor box does not necessarily have to be predetermined. YOLOv8 adopts the dynamic allocation approach of TaskAlignedAssigner to determine the distribution of positive and negative samples, and the weighted score for classification and regression is utilized to select positive samples. The classification loss adopts VFL, and the positioning loss adopts DFL [16] and CIoU [17].

The use of YOLOv8 for fire detection has yielded favorable outcomes. However, considering the significance of fire detection, this research aims to improve the performance of the YOLOv8 model to increase its accuracy in the following ways.

- The newly designed DGIConv module replaces the original Conv module to enhance the performance of the model.
- A newly designed attention mechanism called FourBranchAttention is employed in the YOLOv8 backbone, and a comparison is made with other attention mechanisms to find the best mechanism.
- The loss function is optimized by adopting better hyperparameters of the TaskAlignedAssigner and employing the newly designed GSIoU as an alternative to the CIoU.

This paper first provides a detailed introduction to the design principles and specific content of DGIConv, FourBranchAttention and the new loss function, followed by a comparative experiment. The experiment is divided into four parts. The first part is the attention mechanism comparison experiment, which compares the newly designed FourBranchAttention with several commonly used attention mechanisms to find the best one. The second part is the experiment on hyperparameter selection. The third part involves three improved combination ablation experiments to obtain the most effective combination. The fourth part is a comparison with other popular models, such as SSD/YOLOv7. Finally, the conclusion section is presented.

2. Material and Methods

This study explores the potential improvement of YOLOv8 by replacing the original Conv module, incorporating a newly designed attention mechanism, and optimizing the loss function. These modifications are intended to increase the performance of the model for fire detection.

2.1. DGIConv Module

The newly designed DGIConv convolution module comprises three parts: dilated convolution [18, 19], grouped convolution [20], and an inception module [21, 22]. Dilated convolution introduces a new parameter called the dilation rate to the convolution layer, which defines the interval of each value when the convolution kernel processes the feature map. The benefit is that it maintains the original height and width of the input feature map while increasing the receptive field. Grouped convolution is used to solve the problem of insufficient video memory and is now widely used in various lightweight models to reduce the number of computations and parameters. Inception was employed, and the advantage of utilizing convolution kernels of varying sizes is that it enhances the network's ability to adapt to targets of different sizes in feature maps.

The DGIConv module shown in Figure 1 replaces the original Conv module of YOLOv8. Eight concurrent operations are executed for an input tensor, and the convolution kernel, dilating rate, stride size, and group of each branch are shown in the figure.

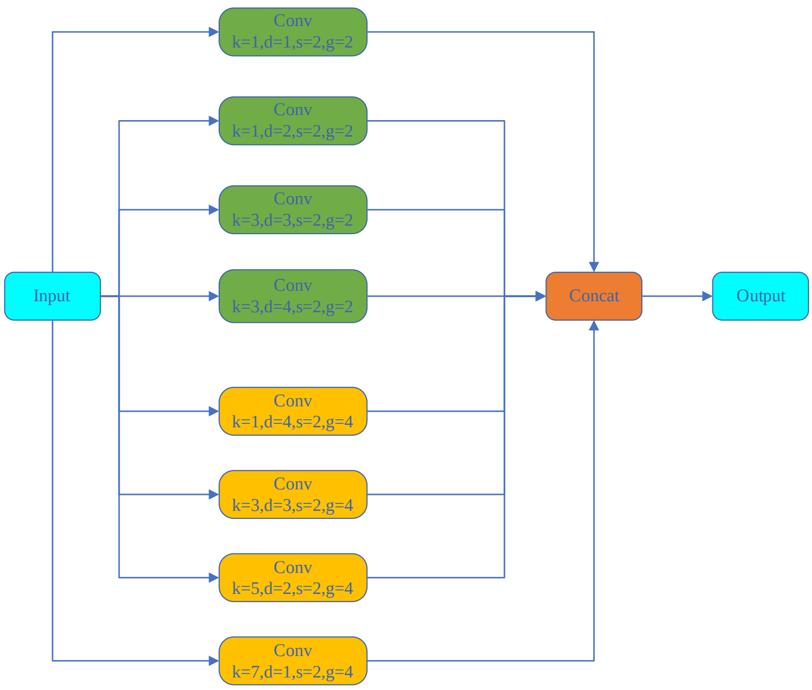

Figure 1. DGIConv structure

The output channel of each branch is set to one-eighth of the corresponding channel in YOLOv8, and the output is obtained by concatenating the results from all branches. Small convolution kernels (3×3 or 1×1) are effective at capturing fine-grained, local features, making them suitable for detecting details and localized information within an image, whereas large kernels (5×5 or 7×7) capture broader contextual information, allowing the network to understand the global structure and large-scale features of a fire image. Different dilation rates allow the network to capture features at multiple scales simultaneously. Smaller dilation rates focus on local details, whereas larger dilation rates capture broader context. This multiscale feature extraction is crucial for fire detection, which requires an understanding of both fine details and global structures. The 2 or 4 groups of convolutions divide the input channels into smaller groups and apply convolutions separately within each group. This reduces the number of computations and parameters, making the model more efficient. Figure 2 illustrates the specific locations of DGIConv in the YOLOv8 backbone network.

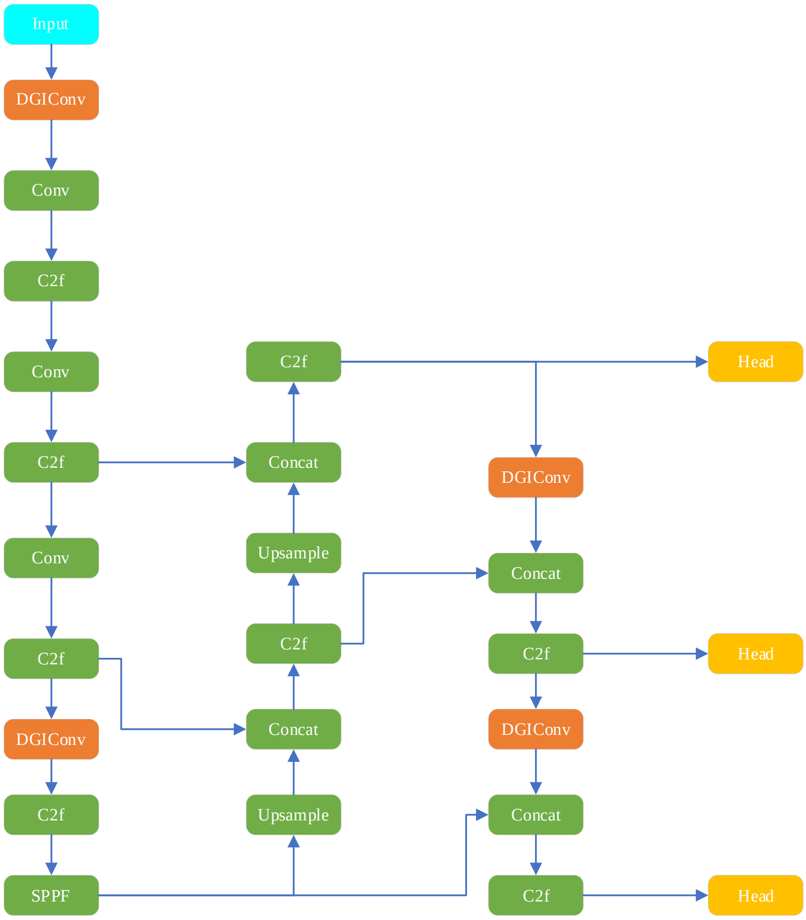

Figure 2. The positions of DGIConv in YOLOv8

2.2. FourBranchAttention

Attention is a mechanism for adjusting parameters to emphasize specific parts within feature maps [23]. The network automatically learns a set of weight coefficients and dynamically adjusts them to amplify the regions of interest while suppressing irrelevant parts. The attention mechanism named FourBranchAttention, which is a newly designed mechanism, is employed in the YOLOv8 backbone network.

As shown in Figure 3, FourBranchAttention refers to the design concept of TripletAttention [24]. The four branches are channel, spatial, C/H, and C/W attention. The C/H branch calculates the channel-height relationships of the feature map, whereas the C/W branch calculates the channel-width relationships of the feature map. Four branches are multiplied by the input, and the results are added to obtain the output. FourBranchAttention enhances the feature representation. By focusing on different dimensions (e.g., spatial, height, width, and channel), FourBranchAttention provides a more comprehensive representation of the features. This multidimensional attention improves the network's ability to capture intricate details and complex patterns in the data of fire images. It captures dependencies along multiple dimensions simultaneously. This allows the network to better understand the spatial and channelwise relationships within the feature maps. Therefore, it can better recognize flame images of different sizes and shapes.

FourBranchAttention mechanisms are compared with four other commonly used attention mechanisms, such as CBAM [25], CoordAttention [26], ShuffleAttention [27], and ECA [28], to find the best one. The attention module incorporated in YOLOv8 is shown in Figure 4 (FBA stands for FourBranchAttention) with four positions, and all the models adopt this structure.

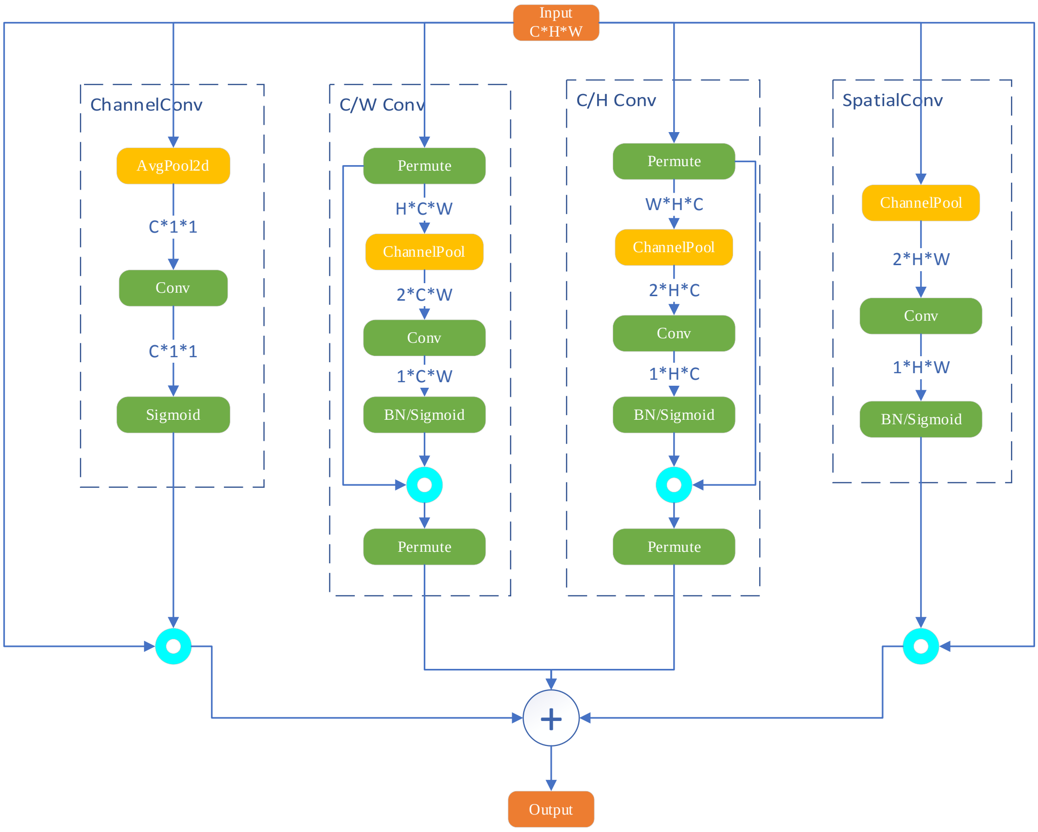

Figure 3. FourBranchAttention structure

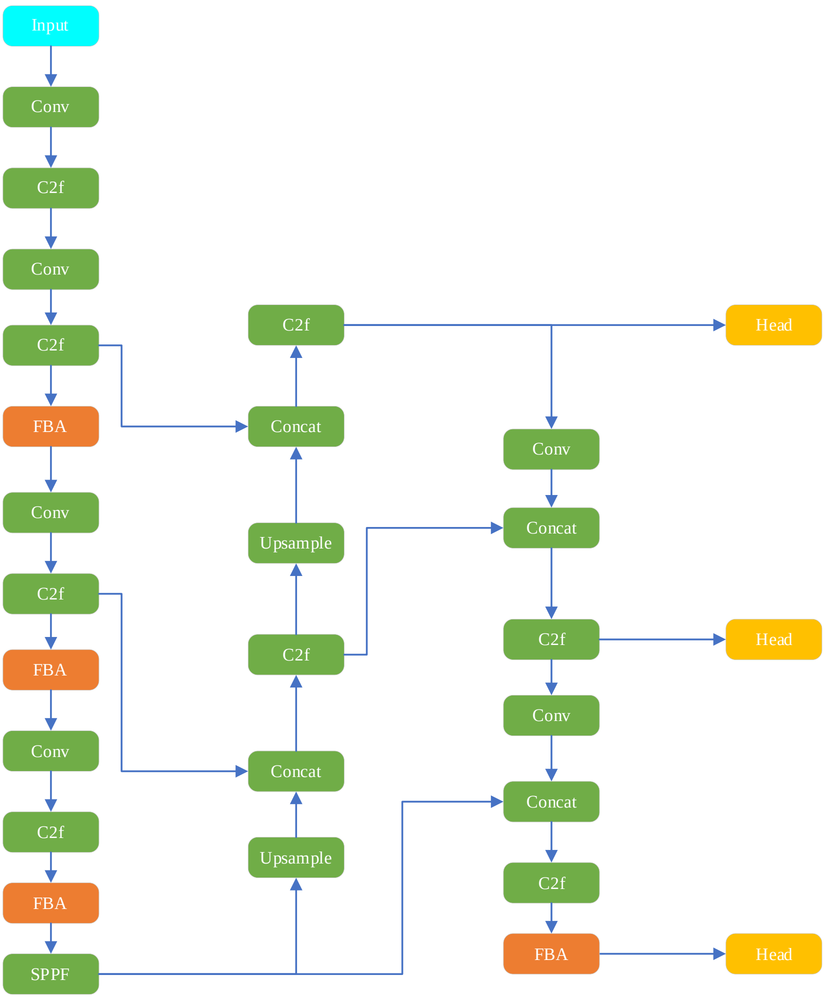

Figure 4. The positions of FourBranchAttention in YOLOv8

2.3. Loss Function Optimization

The loss function of YOLOv8 includes both positive and negative sample allocation strategies and loss calculations. The YOLOv8 algorithm uses the TaskAlignedAssigner dynamic allocation strategy to achieve positive sample matching. The loss calculation comprises two parts: the classification loss VFL, which adopts the cross-entropy algorithm. The localization loss consists of border regression losses, DFL and CIoU. This study improves the hyperparameter selection of the TaskAlignedAssigner and employs the GSIoU as an alternative to the CIoU.

The YOLOv8 algorithm uses the TaskAlignedAssigner dynamic allocation strategy to achieve positive sample matching. Positive samples are selected on the basis of the weighted scores of the classification and regression, as shown in Equation 1, where s represents the classification loss calculated on the basis of the category for all the predicted boxes and where u represents the IoU loss. The TopK prediction boxes are selected as positive samples on the basis of their total scores. TopK, α, and β are hyperparameters set to 10, 0.5, and 6 by default in YOLOv8. We explored the values of these variables to obtain better results.

$$t = s^{\alpha} \times u^{\beta} \tag{1}$$

YOLOv8 uses the CIoU by default, and the CIoU is depicted in Equations 2-4. The IoU stands for the intersection over union. The value of $\rho^2(b^{pd}\ b^{gt})$ corresponds to the Euclidean distance between the center points of the predicted and target boxes, and the value of c reflects the length of a diagonal line of the minimum enclosing rectangle that encompasses both the predicted and target boxes. α is the weight factor, and v is used to measure the consistency of the aspect ratio. w^{gt} and h^{gt} represent the width and height of the ground truth box, respectively. w^{pb} and h^{pb} represent the width and height of the prediction box, respectively. The CIoU incorporates the distance between center points, the IoU area, and the factors in the aspect ratio.

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

$$\alpha = \frac{v}{1 - IoU + v} \tag{3}$$

$$v = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w^{pb}}{h^{pb}}\right) \tag{4}$$

However, if the aspect ratios of the predicted and target boxes are the same, then the penalty term of the aspect ratio is constant at 0 in the CIoU, which is unreasonable. The GSIoU shown in Equations 5-8 adjusts the loss of the IoU via a coefficient that conforms to a Gaussian distribution. To increase the positioning accuracy of the model through an adjustment factor λ related to the IoU, the loss coefficients of low-quality and high-quality anchor boxes are relatively reduced. This allows GSIoU to focus on anchor boxes of mean quality and improves the overall performance of the detector. L_{IoU} represents the IoU loss. R_{GSIoU} represents the distance loss between the ground truth box and the predicted box. The value of $\rho^2(b^{pd}\ b^{gt})$ corresponds to the Euclidean distance between the center points of the predicted and target boxes, and the value of c reflects the length of a diagonal line of the minimum enclosing rectangle that encompasses both the predicted and target boxes. The hyperparameters β, μ, and σ are adjustment factors that are assigned values of 2, 1.5, and 0.75, respectively. L_{IoUn} represents the loss of the IoU during the Nth iteration. L_{IoUn-1} represents the mean loss of the IoU during the N-1 iteration.

$$L_{GSIoU} = \lambda R_{GSIoU} L_{IoU} \tag{5}$$

$$R_{GSIoU} = \frac{\rho^2(b^{pb}, b^{gt})}{c^2} \tag{6}$$

$$\lambda = \beta \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{7}$$

$$x = {L_{IoU}}_n / {L_{IoU}}^*_{n-1} \tag{8}$$

This adjustment factor λ is a function that follows a normal distribution, with high middle and low ends. For the medium-quality prediction boxes with the highest proportion, the addition of this adjustment factor relatively increases the weight of the loss, making them converge quickly and ultimately becoming high-quality boxes. For low-quality prediction boxes, in subsequent iterations, as the medium-quality prediction boxes improve, more low-quality boxes are included in the medium-quality boxes, so they also converge quickly. For an increasing number of high-quality boxes, the loss weight is relatively reduced, but it has reached a good matching level, so there is no need for more adjustments, saving computing resources. The CIoU evenly distributes computing resources, whereas the GSIoU can provide different levels of adjustment to prediction boxes of different qualities. Thus, it can converge faster and better, achieving better results.

3. Results

Currently, there are a limited number of publicly accessible fire datasets on a significant scale. This study incorporates various internet datasets, including the ImageNet and BoWFire datasets. This study gathered and annotated approximately 3000 images (the training and testing sets were divided into 7-3 parts). The optimizer utilized stochastic gradient descent (SGD) with an initial learning rate of 0.01, momentum value of 0.937, and batch size of 32. The experimental setup included an RTX 4060Ti graphics card, Windows 11, PyTorch 1.13.1, CUDA 11.7, and Python 3.8. All the experiments were performed over 200 epochs. The experimental indicators encompass the number of model parameters and computational complexity represented by the GFLOPs. The metric mAP50 denotes the accuracy at an IoU threshold of 0.5, and mAP50-95 indicates the mean accuracy within an IoU range of 0.5-0.95.

3.1. Results of Incorporating Attention Mechanism

Table 1 displays the outcomes of the attention experiments, which include five distinct types of attention: CBAM, CoordAttention (CA), ShuffleAttention (SA), ECA, and FourBranchAttention (FBA). YOLOv8n incorporating FBA yielded the most effective outcomes, with notable improvements of 1.18% in terms of mAP50 and 1.92% in terms of mAP50-95 compared with the baseline. My personal speculation is that FourBranchAttention fully explores the relationships between various dimensions, including channel, spatial, C/H, and C/W attention, surpassing other attention dimensions, thus achieving the best results.

Table 1. Results of incorporating attention mechanisms.

Experiment	Parameters (MB)	GFLOPs	mAP50(%)	mAP50-95(%)
YOLOv8n	3.01	8.20	95.94	71.89
+CBAM	3.16	8.30	96.32	73.23
+CA	3.03	8.20	97.02	73.47
+SA	3.01	8.20	96.68	72.70
+ECA	3.01	8.30	96.33	72.91
+FBA	3.16	8.50	97.12	73.81

3.2. Hyperparameter Selection of TaskAlignedAssigner

YOLOv8 uses a combination of the classification score and IoU loss to measure the overall score of task alignment, where s and u are the classification score and IoU value, respectively, and α and β are the weight hyperparameters. The TopK prediction boxes are selected as positive samples, and the remaining prediction boxes are selected as negative samples on the basis of their scores. In the YOLOv8 model, the default values for α, β, and TopK are 0.5, 6.0 and 10, respectively. However, for the specific task of fire detection, this study further explores the values of these variables to achieve better results.

First, this study fixed α/β and tested the influence of the TopK value on the results. The results are shown in Figures 5-6. When TopK was 9, mAP50 and mAP50-95 achieved the maximum value simultaneously when the experiment reached 30 epochs.

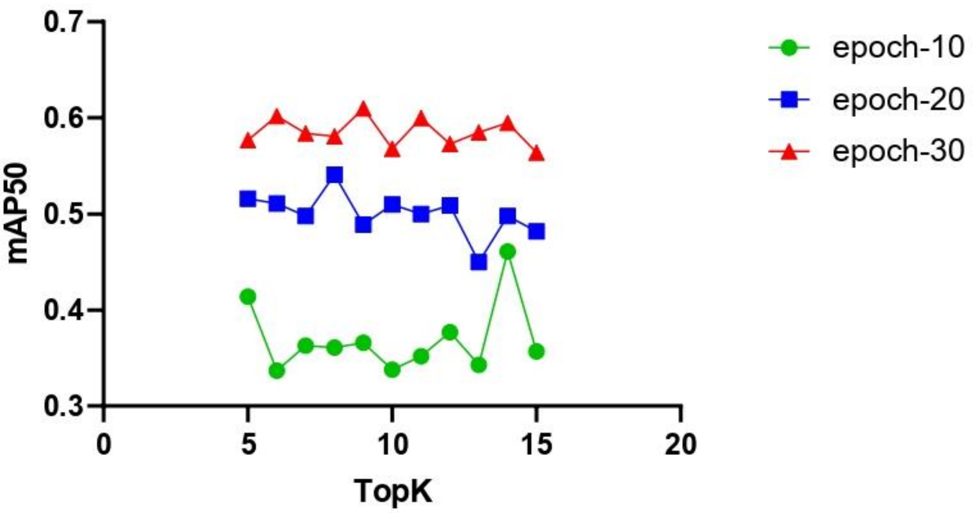

Figure 5. The impact of TopK on the mAP50 values

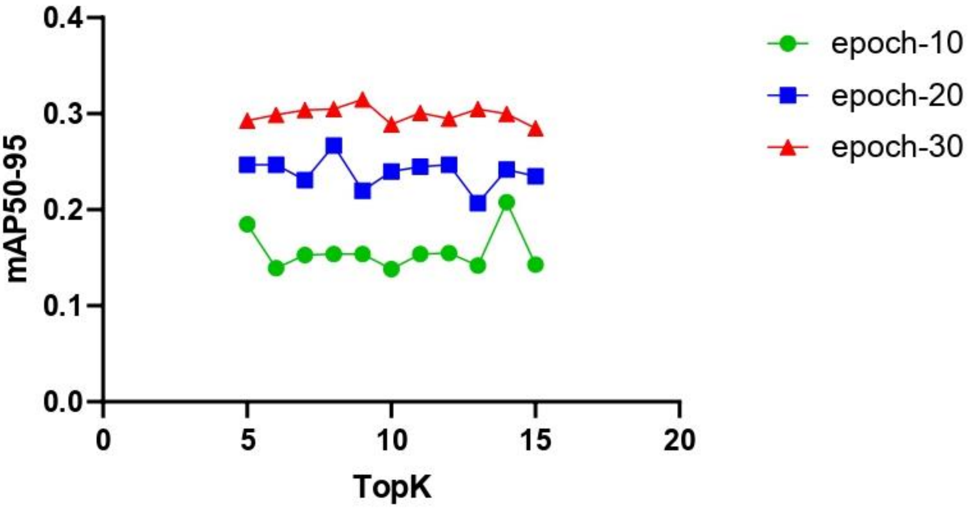

Figure 6. The impact of TopK on the mAP50-95 values

Then, we fixed the TopK value and performed an opposite transformation on the two parameters α and β, which conform to a linear substitution relation, as shown in Equation 9. β ranges from 10 to 1 with a stride of 1, which corresponds to α changing from 0.1 to 1 with a stride of 0.1. The results are shown in Figures 7-8. The comprehensive maximum values of mAP50 and mAP50-95 are obtained when α is 0.6 and β is 5.

$$11 = 10\alpha + \beta \tag{9}$$

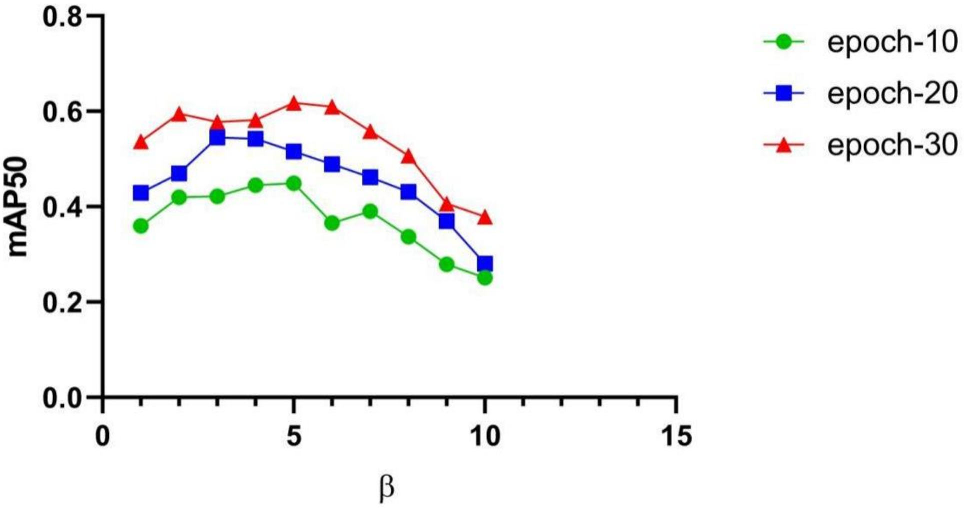

Figure 7. The impact of α and β on the mAP50 values

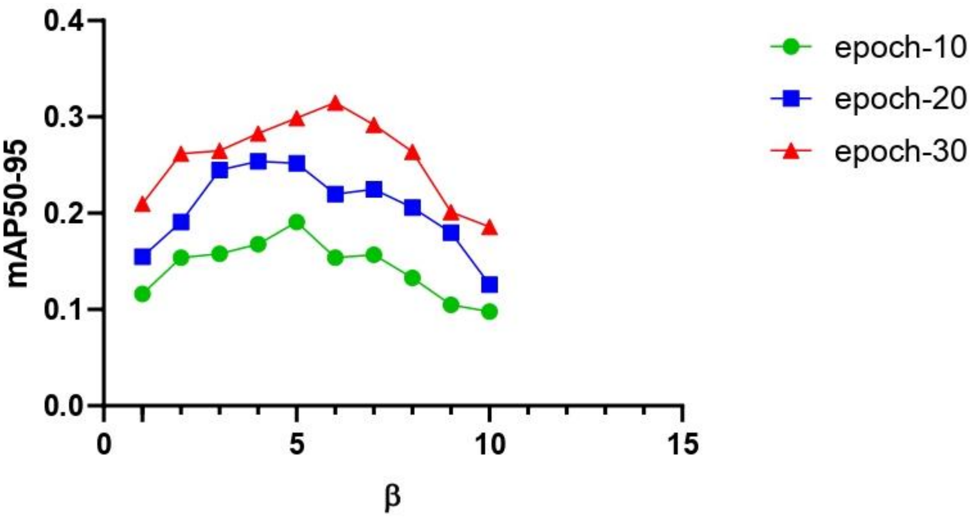

Figure 8. The impact of α and β on the mAP50-95 values

3.3. Ablation Experiment

The DGIConv module, FourBranchAttention, and optimized loss function are employed, and the relevant ablation experimental results are presented in Table 2. DGIConv is used to improve the main structure of the network, FourBranchAttention is used to add attention modules, and the optimized loss function and better hyperparameters are adopted. All the models improved by different methods achieved varying degrees of improvement. The three improvements do not overlap with each other, so the enhanced YOLOv8n with three improvements achieves the best performance in fire detection. Compared with the original YOLOv8n, the enhanced YOLOv8 achieves a reduction in the number of GFLOPs of 1.2% and a growth in the number of parameters of 15.0%; however, there is a notable increase of 2.52% to 98.46% in the mAP50 and 3.37% to 75.26% in the mAP50-95. The curves for mAP50 and mAP50-95 are shown in Figures 9-10.

Table 2. Results of the ablation experiment.

Experiment	Parameters (MB)	GFLOPs	mAP50(%)	mAP50-95(%)
YOLOv8n	3.01	8.20	95.94	71.89
+DGIConv	3.30	7.80	96.77	73.83
+FBA	3.16	8.50	97.12	73.81
+Loss	3.01	8.20	97.57	71.46
+DGIConv+FBA	3.46	8.10	96.69	74.76
+DGIConv+Loss	3.30	7.80	97.62	72.99
+FBA+Loss	3.16	8.50	97.94	75.02
+DGIConv+FBA+Loss	3.46	8.10	98.46	75.26

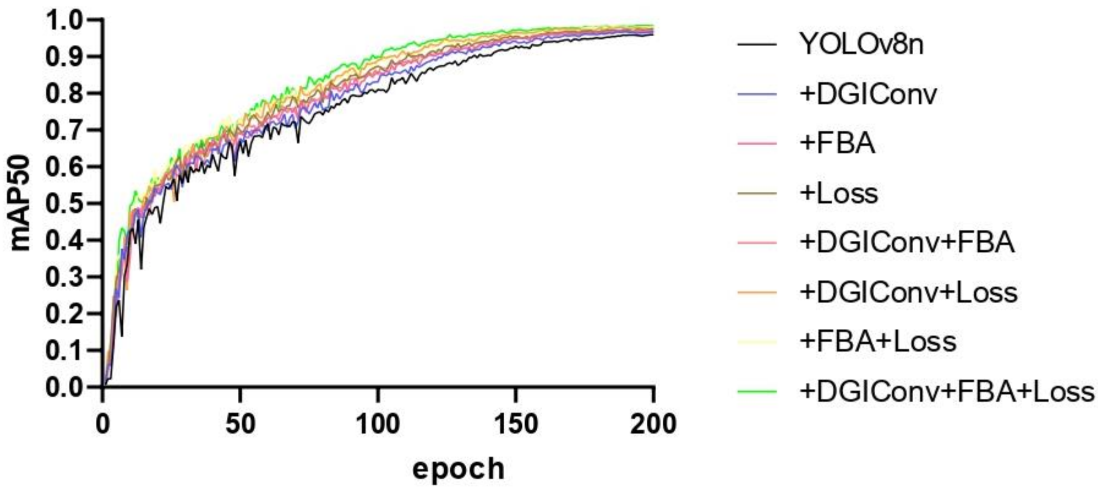

Figure 9. The mAP50 values in the ablation experiment.

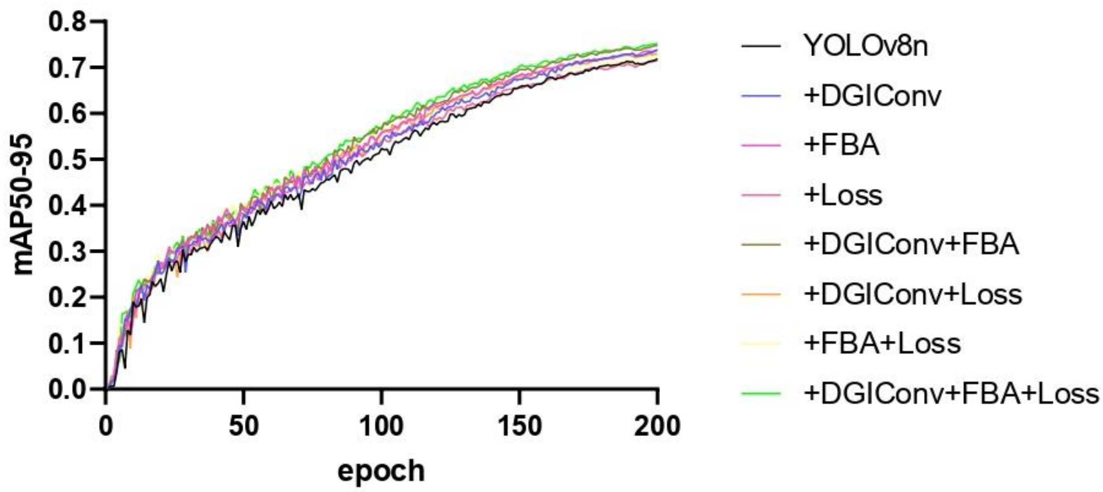

Figure 10. The mAP50-95 values in the ablation experiment.

3.4. Comparison with Other Models

The experimental results of the comparison with other models, including Faster R-CNN, SSD, and YOLOv7, are presented in Table 3. The enhanced YOLOv8n model outperforms Faster R-CNN and SSD with reduced parameter and computational complexity but yields much higher mAP50 and mAP50-95 scores. Compared with YOLOv7-tiny, it significantly reduces the number of parameters by 42.43% and the computational complexity by 38.64%, but it improves the mAP50 by 8.13% and the mAP50-95 by 19.8%. My personal speculation is that YOLOv8 architecture is a single-stage object detector, meaning it predicts bounding boxes and class probabilities directly from full images in a single evaluation. This makes it significantly faster and more efficient than two-stage detectors. Faster R-CNN is a two-stage detector. The first stage generates region proposals, and the second stage classifies these proposals and refines their bounding boxes. This added complexity usually makes it relatively poorer than YOLOv8. SSD is a single-stage detector, but it uses multiple feature maps of different scales, which can be less efficient than YOLOv8's approach.

Table 3. Comparison with other models

Experiment	Parameters (MB)	GFLOPs	mAP50(%)	mAP50-95(%)
FasterR-CNN (resnet50)	28.28	470.46	60.56	28.58
SSD (mobilenet)	3.54	1.40	58.01	27.23
SSD (vgg16)	23.61	60.76	64.21	33.29
YOLOv7	37.20	105.10	97.93	74.55
YOLOv7-tiny	6.01	13.20	90.33	55.46
Enhanced YOLOv8	3.46	8.10	98.46	75.26

4. Conclusions and Discussion

Compared with related studies, this study surpasses them in indicators such as mAP50/mAP50-95, but there are shortcomings in this study, such as a lack of image processing speed. Luo et al. [29] proposed an improved YOLOX target detection algorithm that combined the Swin transformer architecture, the CBAM attention mechanism, and a Slim Neck structure for flame smoke detection in laboratory fires. The experimental results verify that the improved YOLOX algorithm has higher detection accuracy and more accurate position recognition for flame smoke in complex situations, with mAP50 of 92.78% and 92.46% for flame and smoke, respectively, on self-built datasets including 3046 flame targets and 2532 smoke targets. The mAP50 of flame detection in this study is 98.46%, exceeding 92.78% in the aforementioned paper, but it includes smoke, which is lacking in this study. Geng et al. [30] proposed an improved fire and smoke object detection algorithm YOLOFM based on YOLOv5n. The study constructed a FocalNext network. An FPN network that effectively reduces redundant calculations, was also proposed. A new compression decoupled head, named NADH, was also created to enhance the correlation between the decoupling head structure and the calculation logic of the loss function. YOLOFM improved the baseline network's mAP50 and mAP50-95 by 2.2% and 7.9% to 96.3% and 68.8%, respectively. The paper focuses mainly on adjusting the network structure. This study focuses on the adjustment of the attention mechanism and loss function, each with its own emphasis. Yun et al. [31] proposed the FFYOLO model for forest fire detection. The CPDA attention mechanism is designed to enhance the feature extraction capabilities of fire and smoke. It replaces the original detection head with MCDH and introduces GSConv to reduce the number of parameters and complexity while maintaining accuracy. The MPDIoU is introduced to reduce the false and missed detection rates, minimizing the risk of overfitting. FFYOLO achieved an mAP50 of 88.8% and an FPS of 188. Although this study surpasses it in terms of the mAP50 metric, it does not calculate and compare the FPS, which is a defect.

In this study, we adopted the YOLOv8 algorithm for fire detection and improved the accuracy of the model. The improvements involve the integration of the DGIConv module, incorporating FourBranchAttention, and employing the GSIoU and new hyperparameters. The newly designed DGIConv convolution module comprises three parts: dilated convolution, grouped convolution, and an inception module, which are used to enhance the feature extraction capability of the model and reduce its computational complexity. This study has adopted the FourBranchAttention mechanism to enhance the recognition ability of flame targets. The use of the GSIoU loss function has improved the convergence speed and localization ability of the model. Through ablation experiments, the enhanced YOLOv8 with DGIConv, FourBranchAttention, and the optimized loss function demonstrated a 2.52% increase to 98.46% in mAP50 and a 3.37% increase to 75.26% in mAP50-95 compared with the original YOLOv8. Compared with other models, such as Faster R-CNN, SSD, and YOLOv7, the performance indicators are significantly improved, resulting in enhanced accuracy in fire detection. As shown in Figure 11, RCNN, SSD, and YOLOv7 have some missed detections or more redundant frames than YOLOv8 and enhanced YOLOv8 does, whereas enhanced YOLOv8 has fewer missed flames than YOLOv8 does. When 100 fire images were randomly selected for testing, RCNN or SSD generated more redundant prediction boxes than did enhanced YOLOv8, which is approximately equivalent to 20% of the total number of real boxes. The enhanced YOLOv8 algorithm exhibited the best performance in these experiments. In addition, this study annotated and disclosed a new fire dataset, including training and testing sets, with approximately 3000 images.

Figure 11. Comparison of the effects of five algorithms

5. Declarations

5.1. Data Availability Statement

The data presented in this study are available in the article.

5.2. Funding

This study was funded by the Innovation and Entrepreneurship Training Program for College Students of Henan University (number: 20231012002).

5.3. Acknowledgements

I am grateful for the strong support of Henan University's Innovation and Entrepreneurship Training Program.

5.4. Institutional Review Board Statement

Not applicable.

5.5. Informed Consent Statement

Not applicable.

5.6. Declaration of Competing Interest

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6. References

[1] Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision, 2015 International Conference on Computer Vision, ICCV 2015, 1440–1448. doi:10.1109/ICCV.2015.169.

[2] Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2021). A Review of Yolo Algorithm Developments. Procedia Computer Science, 199, 1066–1073. doi:10.1016/j.procs.2022.01.135.

[3] Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: challenges, architectural successors, datasets and applications. Multimedia Tools and Applications, 82(6), 9243–9275. doi:10.1007/s11042-022-13644-y.

[4] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 9905 LNCS, 21–37. doi:10.1007/978-3-319-46448-0_2.

[5] Han, Y., Duan, B., Guan, R., Yang, G., & Zhen, Z. (2024). LUFFD-YOLO: A Lightweight Model for UAV Remote Sensing Forest Fire Detection Based on Attention Mechanism and Multi-Level Feature Fusion. Remote Sensing, 16(12), 2177. doi:10.3390/rs16122177.

[6] Cao, L., Shen, Z., & Xu, S. (2024). Efficient forest fire detection based on an improved YOLO model. Visual Intelligence, 2(1), 20. doi:10.1007/s44267-024-00053-y.

[7] Huang, J., Zhou, J., Yang, H., Liu, Y., & Liu, H. (2023). A Small-Target Forest Fire Smoke Detection Model Based on Deformable Transformer for End-to-End Object Detection. Forests, 14(1), 162. doi:10.3390/f14010162.

[8] Wang, D., Qian, Y., Lu, J., Wang, P., Yang, D., & Yan, T. (2024). EA-YOLO: Efficient Extraction and Aggregation Mechanism of YOLO for Fire Detection, 22. doi:10.21203/rs.3.rs-3930713/v1.

[9] Wei, Z. (2023). Fire Detection of yolov8 Model based on Integrated SE Attention Mechanism. Frontiers in Computing and Intelligent Systems, 4(3), 28–30. doi:10.54097/fcis.v4i3.10765.

[10] Saydirasulovich, S. N., Mukhiddinov, M., Djuraev, O., Abdusalomov, A., & Cho, Y. I. (2023). An Improved Wildfire Smoke Detection Based on YOLOv8 and UAV Images. Sensors (Basel, Switzerland), 23(20), 8374. doi:10.3390/s23208374.

[11] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December, 779–788. doi:10.1109/CVPR.2016.91.

[12] Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. arXiv, Preprint arXiv:1811.03378.

[13] Jiang, C., Ren, H., Ye, X., Zhu, J., Zeng, H., Nan, Y., Sun, M., Ren, X., & Huo, H. (2022). Object detection from UAV thermal infrared images and videos using YOLO models. International Journal of Applied Earth Observation and Geoinformation, 112, 102912. doi:10.1016/j.jag.2022.102912.

[14] Tian, Z., Shen, C., Chen, H., & He, T. (2022). FCOS: A Simple and Strong Anchor-Free Object Detector. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(4), 1922–1933. doi:10.1109/TPAMI.2020.3032166.

[15] Zhang, S., Chi, C., Yao, Y., Lei, Z., & Li, S. Z. (2020). Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 9756–9765. doi:10.1109/CVPR42600.2020.00978.

[16] Zhou, Y., Zhu, W., He, Y., & Li, Y. (2023). YOLOv8-based Spatial Target Part Recognition. Proceedings of 2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence, ICIBA 2023, 3, 1684–1687. doi:10.1109/ICIBA56860.2023.10165260.

[17] Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-IoU loss: Faster and better learning for bounding box regression. AAAI 2020 - 34th AAAI Conference on Artificial Intelligence, 34(07), 12993–13000. doi:10.1609/aaai.v34i07.6999.

[18] Yu, F., & Koltun, V. (2016). Multi-scale context aggregation by dilated convolutions. In 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings.

[19] Wei, Y., Xiao, H., Shi, H., Jie, Z., Feng, J., & Huang, T. S. (2018). Revisiting Dilated Convolution: A Simple Approach for Weakly- and Semi-Supervised Semantic Segmentation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 7268–7277. doi:10.1109/CVPR.2018.00759.

[20] Chen, T., Duan, B., Sun, Q., Zhang, M., Li, G., Geng, H., Zhang, Q., & Yu, B. (2022). An Efficient Sharing Grouped Convolution via Bayesian Learning. IEEE Transactions on Neural Networks and Learning Systems, 33(12), 7367–7379. doi:10.1109/TNNLS.2021.3084900.

[21] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 31(1), 4278–4284. doi:10.1609/aaai.v31i1.11231.

[22] Nandini, B. (2021). Detection of Skin Cancer using Inception V3 And Inception V4 Convolutional Neural Network (CNN) For Accuracy Improvement. Revista Gestão Inovação e Tecnologias, 11(4), 1138–1148. doi:10.47059/revistageintec.v11i4.2174.

[23] Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. Neurocomputing, 452, 48–62. doi:10.1016/j.neucom.2021.03.091.

[24] Misra, D., Nalamada, T., Arasanipalai, A. U., & Hou, Q. (2021). Rotate to attend: Convolutional triplet attention module. Proceedings - 2021 IEEE Winter Conference on Applications of Computer Vision, WACV 2021, 3138–3147. doi:10.1109/WACV48630.2021.00318.

[25] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 11211 LNCS, 3–19. doi:10.1007/978-3-030-01234-2_1.

[26] Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 13708–13717. doi:10.1109/CVPR46437.2021.01350.

[27] Zhang, Q. L., & Yang, Y. Bin. (2021). SA-Net: Shuffle attention for deep convolutional neural networks. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2021-June, 2235–2239. doi:10.1109/ICASSP39728.2021.9414568.

[28] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 11531–11539. doi:10.1109/CVPR42600.2020.01155.

[29] Luo, M., Xu, L., Yang, Y., Cao, M., & Yang, J. (2022). Laboratory Flame Smoke Detection Based on an Improved YOLOX Algorithm. Applied Sciences (Switzerland), 12(24), 12876. doi:10.3390/app122412876.

[30] Geng, X., Su, Y., Cao, X., Li, H., & Liu, L. (2024). YOLOFM: an improved fire and smoke object detection algorithm based on YOLOv5n. Scientific Reports, 14(1), 4543. doi:10.1038/s41598-024-55232-0.

[31] Yun, B., Zheng, Y., Lin, Z., & Li, T. (2024). FFYOLO: A Lightweight Forest Fire Detection Model Based on YOLOv8. Fire, 7(3), 93. doi:10.3390/fire7030093.