

Article

Precision-Boosted Forest Fire Target Detection via Enhanced YOLOv8 Model

Zhaoxu Yang¹, Yifan Shao¹, Ye Wei¹ and Jun Li^{2,*}

¹ School of Computer and Communication Technology, Lanzhou University of Technology, Lanzhou 730050, China; 210162901045@lut.edu.cn (Z.Y.); 210031501038@lut.edu.cn (Y.S.); 220010501048@lut.edu.cn (Y.W.)

² Department of Applied Mathematics, Lanzhou University of Technology, Lanzhou 730050, China

* Correspondence: junli@lut.edu.cn

Abstract: Forest fires present a significant challenge to ecosystems, particularly due to factors like tree cover that complicate fire detection tasks. While fire detection technologies, like YOLO, are widely used in forest protection, capturing diverse and complex flame features remains challenging. Therefore, we propose an enhanced YOLOv8 multiscale forest fire detection method. This involves adjusting the network structure and integrating Deformable Convolution and SConv modules to better adapt to forest fire complexities. Additionally, we introduce the Coordinate Attention mechanism in the Detection module to more effectively capture feature information and enhance model accuracy. We adopt the WIoU v3 loss function and implement a dynamically non-monotonic mechanism to optimize gradient allocation strategies. Our experimental results demonstrate that our model achieves a mAP of 90.02%, approximately 5.9% higher than the baseline YOLOv8 network. This method significantly improves forest fire detection accuracy, reduces False Positive rates, and demonstrates excellent applicability in real forest fire scenarios.

Keywords: forest fire detection; YOLOv8; SConv; deformable convolution; CoordAtt mechanism; WIoU v3



Citation: Yang, Z.; Shao, Y.; Wei, Y.; Li, J. Precision-Boosted Forest Fire Target Detection via Enhanced YOLOv8 Model. *Appl. Sci.* **2024**, *14*, 2413. <https://doi.org/10.3390/app14062413>

Academic Editor: Sungho Kim

Received: 20 February 2024

Revised: 3 March 2024

Accepted: 5 March 2024

Published: 13 March 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

Forests, the most extensive and biologically diverse terrestrial ecosystems on Earth, play a crucial role in various essential functions, such as water purification, climate regulation, and maintenance of biodiversity [1,2]. Serving as a key pillar of global ecological security, forests are irreplaceable in preserving overall ecological balance and human well-being. With the ongoing evolution of the global climate, the increasing frequency and intensification of forest fires pose escalating challenges to the Earth's ecosystems and human societies [3,4]. These fires not only result in extensive vegetation destruction and habitat loss but also have profound impacts on ecological balance and global climate. In this context, the upgrading of fire monitoring technology becomes an imperative task to enhance the capabilities of early warning, rapid response, and flexible control of forest fires.

The rapid spread of forest fires is attributed to swift air convection and ample oxygen within the forest [5], leading to the rapid escalation of flames within a short timeframe. In such a scenario, early fire detection becomes particularly urgent. While manual inspection was the earliest method, it was swiftly replaced by sensor-based detection systems, due to expensive manpower costs and inefficiency. Sensor technologies [6,7], encompassing smoke sensors, gas sensors, temperature-humidity sensors, etc., exhibit remarkable performance in confined indoor environments. However, they face a series of challenges in large open spaces or extensive areas such as forests, including limited detection distances, high installation costs, and complex communication and power network issues. Moreover, these sensors cannot provide crucial visual information to assist firefighters in rapidly understanding the situation at the fire scene. Although satellite remote sensing can offer

clear images [8], it falls short in achieving real-time detection, especially when affected by weather and cloud cover. In comparison, cameras as sensors possess superior real-time performance, particularly when installed on drones, enabling effective detection in more remote forest areas.

Researchers have proposed a series of highly effective fire detection systems by integrating color characteristics, color-space models, and motion features using various methodologies. Chen, Y. et al. [9] have successfully created a holistic application for forest fire detection and monitoring. This application harnesses flame chromaticity and the Lab color model, incorporating techniques for enhancing image quality and processing. MAI Mahmoud et al. [10] utilized background subtraction and a color-segmentation model, employing support vector machines to classify areas as either genuine fires or non-fires. These studies emphasize the holistic application of flame chromaticity, color space, and motion features, providing diverse and refined approaches to forest fire detection. Prema et al. [11] introduced an effective approach for addressing various fire detection scenarios. This method involves the extraction of texture features, both static and dynamic, from regions identified as potential fire areas, utilizing the YVbCr color model. Additionally, Han et al. [12] focused on flame extraction, facilitating the precise localization of potential fire locations through the integration of multi-color detection and motion points in RGB, HIS, and YUV color spaces.

Recently, there has been a trend towards combining prevalent algorithms in machine learning or deep learning with remote sensing techniques and fire detection systems [13]. Mahaveerakannan R. et al. [14] proposed a Cat Swarm Fractional Calculus Optimization algorithm for deep learning, combining the optimal features of Cat Swarm Optimization with fractional calculus to achieve superior training results. Barmpoutis et al. [15] utilized the Faster R-CNN architecture for fire detection, proposing an approach that integrates both deep learning networks and multidimensional texture analysis to identify potential fire areas through image analysis. K. Alice et al. [16] utilized the AFFD-ASODTL model, incorporating Atom Search Optimizer and deep transfer learning for automated forest fire detection, thereby reducing response time and minimizing wildfire destruction. Ji Lin et al. [17] proposed TCA-YOLO, an efficient and accurate global forest fire detection model based on YOLOv5 and a Transformer encoder. The model includes a coordinated attention mechanism and adaptive spatial feature-fusion technology, reducing manual labeling efforts through semi-supervised learning. TCA-YOLO demonstrates excellent performance across various scenarios. Ghali et al. [18] proposed an innovative ensemble-learning technique that merges the EfficientNet-B5 and DenseNet-201 models, aiming to detect wildfires in aerial images, achieving superior performance compared to various benchmarks in wildfire classification.

Currently, YOLO has made full-size strides in the subject of object detection. However, when confronted with far-off sensing, particularly in the context of wooded-area fires, the mannequin, as is well known, shows sure limitations. Specifically, detection pursuits in wooded-area fires frequently exist in a diminutive state, imparting a unique challenge, particularly in the early ranges of a fire. In some instances, flames and smoke all through a fire site might also be obscured by way of dense foliage, and their normally small scale makes them hard for the mannequin to seize effortlessly. The intrinsic traits of woodland fires pose significant challenges to the overall detection performance of the YOLOv8 model. Consequently, situations of overlooked detection may additionally arise, whereby the mannequin fails to figure without delay the preliminary symptoms of a fire. This no longer solely influences the timeliness of hearth detection but additionally poses an undertaking to the universal accuracy of the network. Despite YOLOv8 showcasing promise in the subject of object detection, its efficacy in the early and far-off detection of woodland fires needs similar refinement and exploration.

To address these challenges, our study presents an enhanced multiscale forest fire detection approach based on YOLOv8. The key contributions are as follows:

- Integration of the Deformable Convolution module [19] into the C2f network structure introduces the *DCN_C2f* module. This integration enables adaptive adjustment of the receptive field, enhancing the model's ability to capture spatial context and object details effectively.
- Creation of the *SCConv_C2f* module by integrating the Spatial Channel-wise Convolution module [20] into the C2f network structure. This aids in optimizing feature representation by capturing channel-wise relationships and semantic information across feature maps.
- Refinement of the Detect module and introduction of the Coordinate Attention mechanism [21], which effectively considers and utilizes both inter-channel relationships and positional information.
- Integration of WIoU v3 [22] into the bounding box regression loss, employing a dynamic non-monotonic mechanism to design a more practical gradient allocation strategy. This reduces gradient acquisition for both high- and low-quality samples, enhancing the model's localization performance and improving its generalization ability.

The remaining Sections of this manuscript comprise the following structure: Section 2 introduces a specifically tailored dataset for forest fire classification, presenting the methodologies and modules utilized in the conducted experiments. This Section further elucidates the proposed model for forest fire classification detection. Section 3 furnishes metrics to assess model performance, showcasing experimental results highlighting improvements in each segment. Section 4 outlines discussions and analyses of the model, along with considerations for future work. Section 5 provides a comprehensive summary of the entire study.

2. Materials and Methods

2.1. Datasets

In the realm of YOLOv8 wildfire detection, the dataset's quality significantly shapes the outcomes of the training process. We utilized a web-scraping approach to gather images relevant to forest fires, encompassing a diverse range of both typical fire scenarios and non-fire environments, for model training. Additionally, we extracted images depicting ground fires, canopy fires, and other large-scale instances exceeding 322 pixels from downloaded wildfire videos. This meticulous approach ensured the dataset's comprehensiveness and diversity, which are essential for effective model training. Furthermore, some researchers have generously shared fire datasets publicly, such as the BoWFireDataset [23].

In total, we amassed 2692 images, consisting of 816 photos depicting woodland-fire situations and 1876 photos depicting fire-free wooded-area environments. Subsequently, this collection was used to compile a diverse dataset focusing on forest fires. To ensure a robust evaluation, we allocated 80% of the dataset for training purposes, reserving 10% for validation and an additional 10% for the test set. It is worth noting that the partitioning process was carried out randomly to maintain the diversity and representativeness of the training and validation sets. Examples of traditional forest fire scenes and aerial imagery captured by drones are depicted in Figure 1.

2.2. YOLOv8

The YOLOv8 algorithm [24], the present day in the YOLO sequence [25], offers 5 vital models: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. In comparison to different algorithms in the YOLO series, its detection concepts resemble these of YOLOv5 [26] and YOLOv7 [27], consisting of 4 major components: input, backbone, neck, and head. Figure 2 illustrates the shape of the YOLOv8 model.



Figure 1. Illustration depicting the structure of the forest fire dataset.

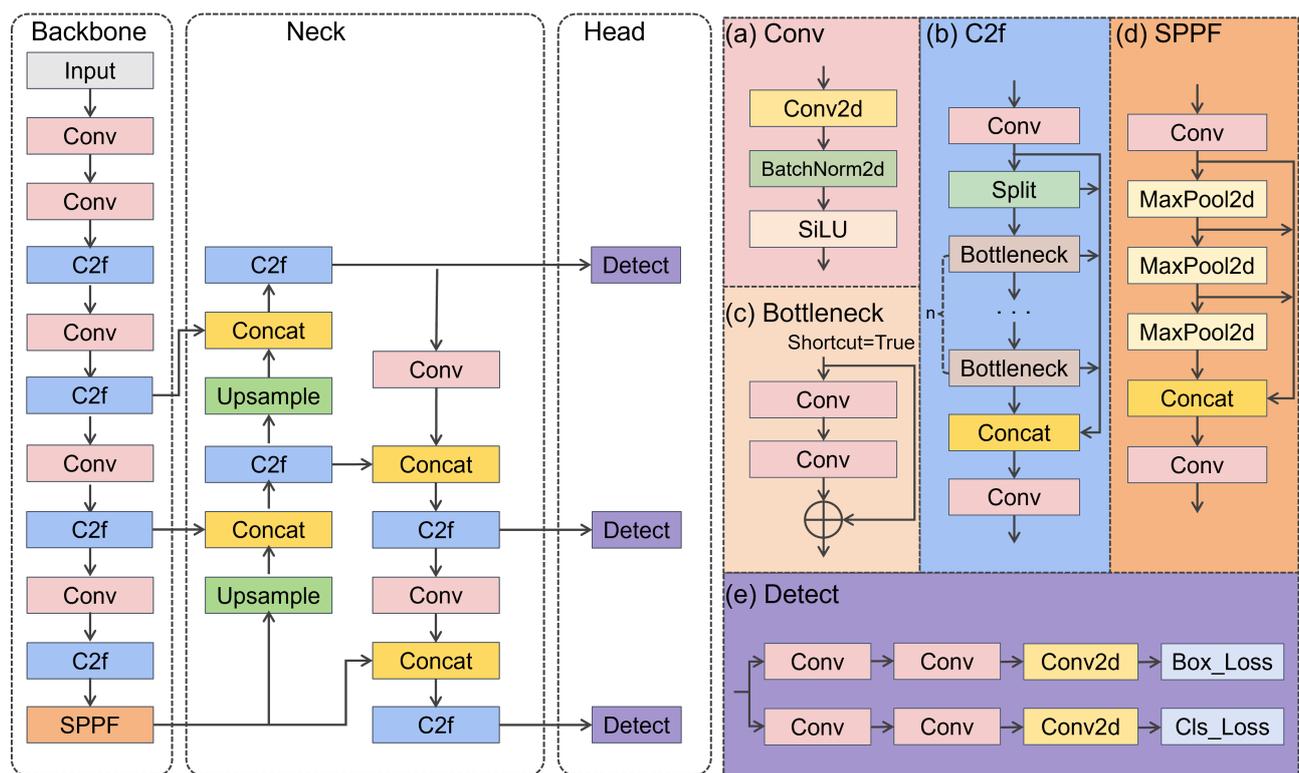


Figure 2. The architecture of YOLOv8.

The input undergoes preprocessing, including data augmentation, before entering the backbone network for feature extraction. The neck module consolidates the extracted features to achieve elements of different sizes (large, medium, small), which are then forwarded to the detection head for the result output. YOLOv8’s backbone network comprises convolutional layers, C2f, and SPPF (Spatial Pyramid Pooling-Fast) modules. The C2f module, derived from YOLOv5’s C3 module and YOLOv7’s ELAN module, preserves the original C3 module’s transition layer shape while modifying the computation blocks. The SPPF module utilizes pyramid pooling to pool the input feature map, generating

sub-feature maps for specific pooling sizes, which are combined to create a large feature map, capturing contextual information at distinct scales. The neck module in YOLOv8 adopts the PAN-FPN (Path Aggregation Network with Feature Pyramid Network) design, enhancing semantic features through characteristic pyramids and FPN cascading down, and propagating finding elements up through PAN. In comparison to YOLOv5, YOLOv8 replaces the PAN-FPN upsampling stage's C3 module with the C2f module, eliminating the CBS 1×1 convolutional structure. It directly integrates characteristics sourced from various levels of the backbone into the upsampling procedure, thereby amplifying the detection prowess for targets exhibiting diverse sizes and shapes. In the Head module, YOLOv8 adopts a decoupled approach, removing the object branch while preserving solely the decoupled classification and regression losses. Furthermore, YOLOv8 transitions from anchor-based to anchor-free, eliminating the need for predefined anchor boxes, making the regression of bounding box loss more challenging.

Regarding training data augmentation, YOLOv8 incorporates the concept proposed by YOLOX [28], disabling mosaic augmentation in the final 10 training rounds. In addition to concat augmentation, it supports MixUp [29] and CopyPaste data augmentation. Mixed data augmentation generates new vectors and labels through random selection and linear interpolation, while CopyPaste data augmentation involves copying portions of an image and pasting them onto other images.

2.3. Improved YOLO8 Model

This section unveils an upgraded model for detecting and categorizing forest fires, utilizing the YOLOv8 framework, as depicted in Figure 3. During the input phase, we initially enhance detection accuracy and discriminative capability through adaptive pre-processing methods, such as image scaling, concatenation, and anchor-box computation. Subsequently, adjustments are made to the C2f network structure, merging it with Deformable Convolution and SCConv to form *DCN_C2f* [30] and *SCConv_C2f*. The middle part of the backbone network's C2f is replaced with *DCN_C2f*, while the remaining parts are substituted with *SCConv_C2f*. Introducing the Deformable Convolution C2f module enables adaptive adjustments to the network's receptive field. Additionally, including the SCConv module not only effectively reduces model parameters and FLOPs but also enhances feature representation. This optimization enables better adaptation to the challenges of forest fire detection and addresses feature redundancy issues in detecting forest fires from unmanned aerial vehicles.

Moreover, the Detect module undergoes improvements to address the directional and position-sensitive characteristics of fires by introducing Coordinate Attention. This attention mechanism concurrently considers inter-channel relationships and positional information, capturing and leveraging these aspects more effectively for precise fire detection. Lastly, in the prediction phase, loss functions are utilized for position calculation, adapting the bounding box loss function to WIoU loss. This adjustment enhances the loss function with directional attributes, fortifying the capabilities of the detection algorithm in training and inference.

2.3.1. C2f Module with Integrated Deformable Convolution Network

In the YOLOv8 backbone architecture, the C2f module improves detection accuracy by leveraging features of diverse scales and integrating contextual information. Nevertheless, stacking kernels may introduce redundant channel information, and employing generic convolutional kernels with fixed characteristics constrains the receptive field, capturing only localized object details. This issue results in potential omissions, particularly when addressing multi-scale, multi-target, or occluded scenarios.

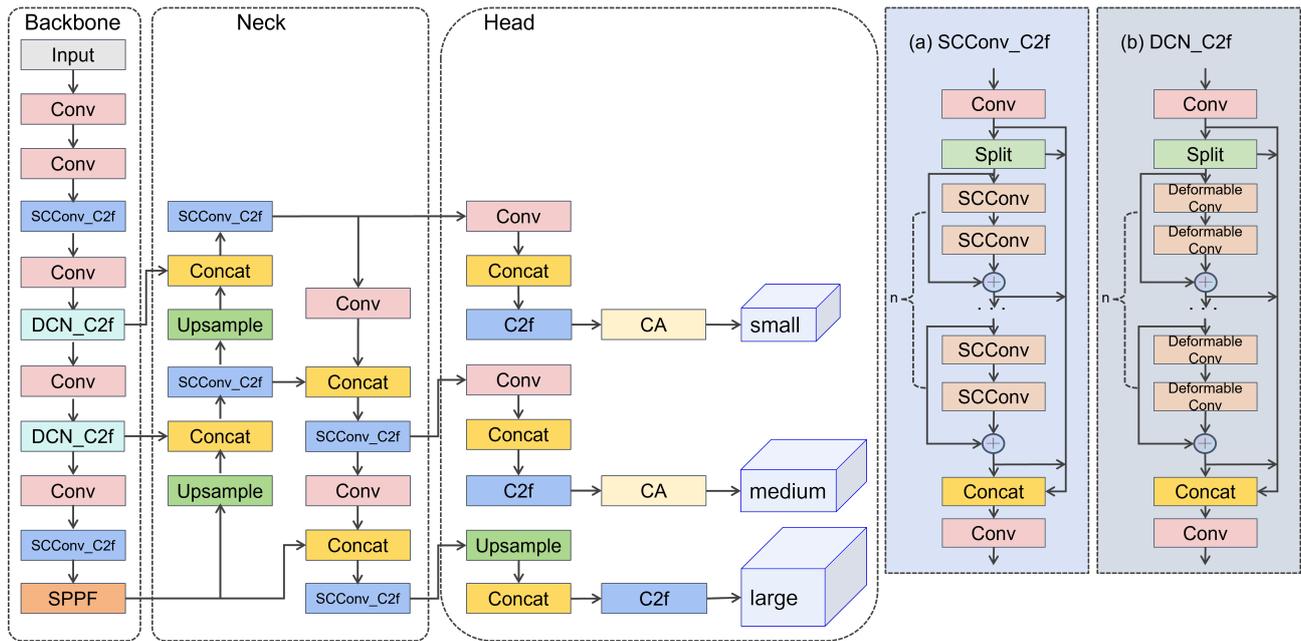


Figure 3. Detailed architecture of the improved YOLO8 mode: small, medium, large denote the detection head with different sizes.

In the context of forest fires, targets such as fire sources and smoke may vary in scale, necessitating precise detection across a wide range from small to large. Conventional object-detection models may encounter difficulties in handling multi-scale targets, particularly in environments where factors like trees and terrain contribute to partial occlusion. Noise or interference can also increase the likelihood of False Positives or misses, especially for smaller objects.

To overcome these challenges, DCN is introduced, employing deformable convolutions in the second and third C2f modules. This adaptation allows the model to better align with the structure and size of objects, improving robustness, especially in regions with small objects. In the process of the model’s prediction and refinement stages, the algorithm adjusts the regression parameters for expected bounding boxes, thus enhancing the model’s effectiveness in handling various target sizes and occlusions. Detailed information about the *DCN_C2f* module is provided in Figure 4.

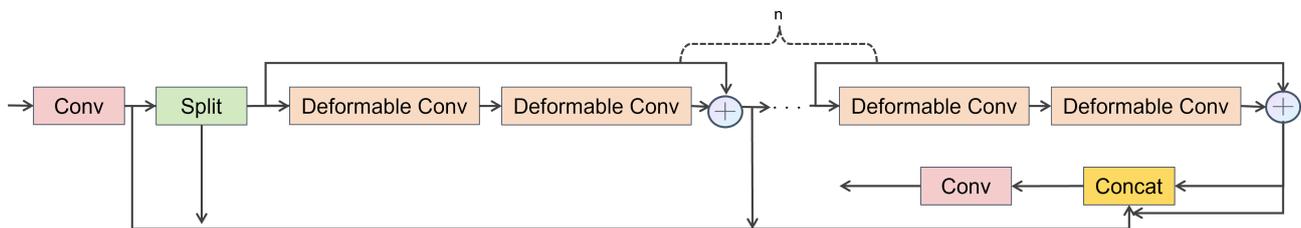


Figure 4. Structure of the *DCN_C2f* module.

The *DCN_C2f* module modifies enter function channels with the usage of a 1×1 convolution, concurrently using the Split operation alternative of a 1×1 convolution for function splitting. By stacking more than one deformable convolution module, the network’s receptive area is expanded, making full use of greater pass connections. This strategy no longer solely reduces the wide variety of parameters but additionally creates a shape wealthy in gradient flow, extracting greater numerous and multi-scale features. DCN enhances the model’s capacity to extract invariant features, permitting convolutional kernels to adapt to the geometric shapes of objects through discovered offsets. By mastering the finest convolutional kernel constructions for one-of-a-kind goal data, DCN improves

the characteristic extraction functionality for infrared aims of exceptional scales. Figure 5 compares general-convolution and deformable-convolution usage of DCN sampling.

In standard convolution, computations are performed on a fixed grid denoted as R , where each sampling factor undergoes weight calculations using the convolutional kernel. In contrast, deformable convolution introduces offsets based totally on well-known convolution. For instance, a 3×3 convolutional kernel with a growth price of 1 can be represented as:

$$R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\},$$

where each element represents the deviation of the convolutional kernel function from the center position.

The output feature matrix of standard convolution at the position p_0 of the central sampling factor is denoted as:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n).$$

In the equation, x denotes the input feature map, while y represents the output feature map. N signifies the total sampling factors, and n enumerates the sampling points; $w(p_n)$ indicates the weight at each position, and $x(p_n)$ signifies the pixel value of the input feature map at that specific location.

After sampling the enter function map x , the function matrix of deformable convolution, delivered with offsets Δp_n (where $n = 1, 2, \dots, N$), is expressed as:

$$y(p_0) = \sum_{p_0 \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n),$$

where Δp_n signifies the offset at the position p_n . As Δp_n typically involves decimal values, $x(p_0 + p_n + \Delta p_n)$ may not correspond directly to the existing points on the feature map. Therefore, DCN utilizes an interpolation method to calculate the offset effect.

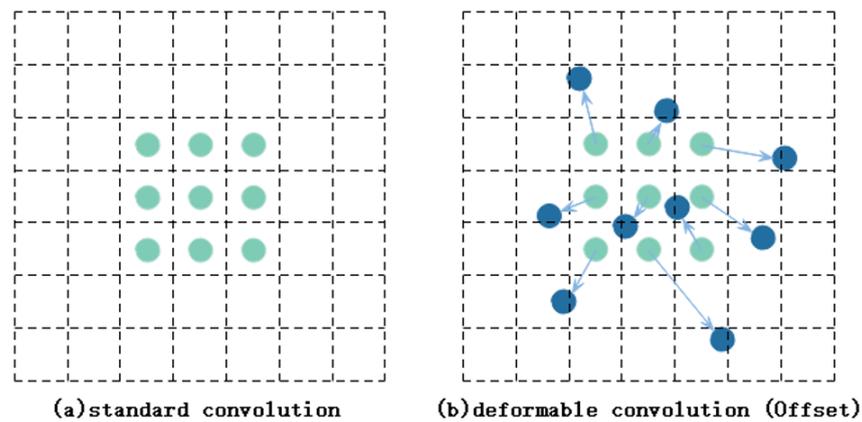


Figure 5. Sampling technique comparison: Standard Convolution vs. Deformable Convolution. The green dots represent the regular sampling grid of standard convolution, while the blue dots indicate the deformed sampling positions with augmented offsets in deformable convolution.

2.3.2. C2f Module with Integrated SCConv

The SCConv module, visually depicted in Figure 6, introduces a groundbreaking CNN compression technique that concurrently addresses spatial and channel redundancy within convolutional layers. This module, proposing two core components, particularly the Spatial Reconstruction Unit and Channel Reconstruction Unit, achieves tremendous overall-performance beneficial properties whilst extensively reducing computational load.

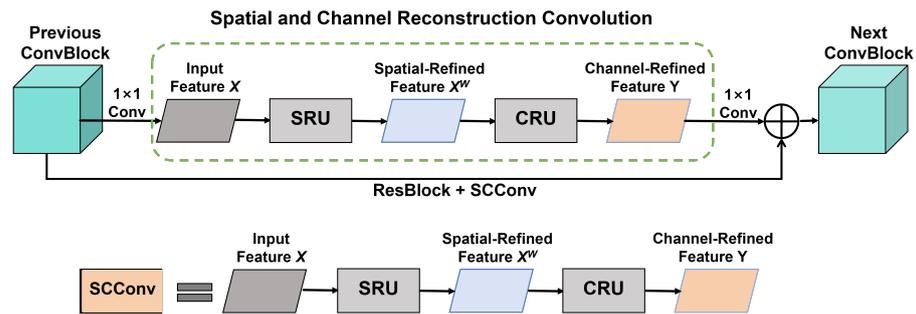


Figure 6. The structure of SCConv is built-in with the SRU and CRU. This diagram indicates the specific function of our SCConv module inside a ResBlock.

SRU addresses spatial-dimension challenges in feature maps by decomposing the entire feature map into multiple spatial blocks and utilizing distinct convolutional kernels for each block, thereby reducing spatial redundancy. This strategy precisely captures characteristic statistics inside every block, appreciably diminishing basic spatial redundancy, bettering function-extraction efficiency, and, at the same time, reducing computational complexity.

On the other hand, CRU focuses on optimizing the channel dimension of characteristic maps, introducing a lightweight wholly linked layer to deal with channel records more flexibly and efficiently. This permits the mannequin to combine records from one-of-a-kind channels more effectively, decreasing channel redundancy. This approach not only strengthens characteristic discrimination but additionally similarly reduces computational requirements.

Functioning as a versatile and compatible component, SCConv seamlessly integrates without necessitating alterations to current model structures, and it can readily substitute standard convolution. In extensive experiments, embedding SCConv into various state-of-the-art methods for image classification and object detection demonstrated superior performance and efficiency balance.

In the context of woodland wildfire detection, SCConv takes the place of the Bottleneck in the original YOLOv8 model’s C2f. This structural alteration transforms C2f into *SCConv_C2f*, effectively streamlining standard convolution and concurrently reducing computational demands. The network architecture for *SCConv_C2f* is illustrated in Figure 7.

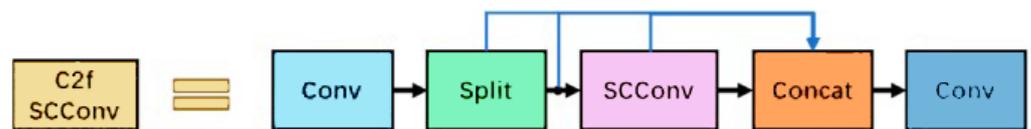


Figure 7. The structure of the *SCConv_C2f* module.

2.3.3. Detection with Integrated CA Attention Mechanism

In the extraction of crucial features from forest fire images with complex backgrounds, YOLOv8 incorporates the Coordinate Attention module. The CA attention module, introduced by Hou Qibin and colleagues, aims to capture extensive dependencies among image features [21]. Contrary to conventional attention mechanisms, Coordinate Attention selectively focuses on spatial positions pertinent to the task at hand. The diagram of CA is depicted in Figure 8.

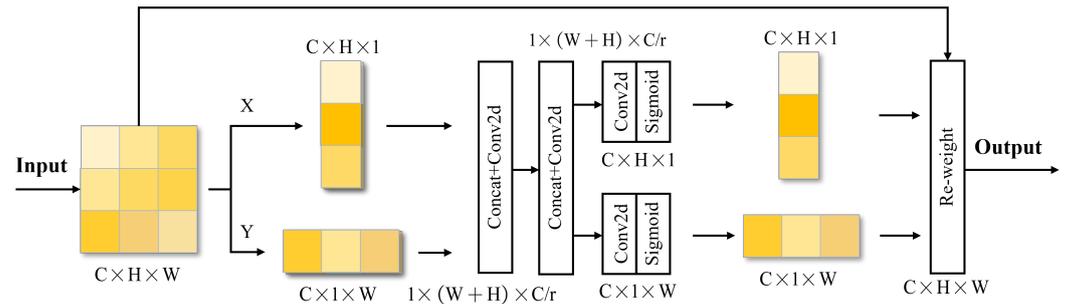


Figure 8. The schematic diagram of CoordAtt.

CoordAtt embeds positional data into channel attention, cleverly fending off the introduction of 2D world pooling. Instead, it employs an approach of decomposing channel interest into two one-dimensional characteristic encodings. This approves the enter elements to be cleverly aggregated into two independently direction-aware function maps alongside the vertical and horizontal directions. These characteristic maps no longer solely comprise embedded directional facts but additionally seize long-distance spatial dependencies alongside the spatial path through the interest maps generated through encoding. Finally, these two interest maps are accelerated with the enter function map to emphasize the expression of the areas of interest.

During the embedding of coordinate information, the challenge of global pooling is addressed by horizontally and vertically decomposing pooling. Specifically, for each feature output, it is represented as:

$$z_{ch}(h) = \frac{1}{W} \sum_{i=0}^{W-1} x_c(h, i),$$

$$z_{cw}(w) = \frac{1}{H} \sum_{j=0}^{H-1} x_c(j, w),$$

where H and W denote the height and width of the pooling kernel, respectively. These operations combine elements across two spatial orientations, generating a set of direction-aware feature maps that capture spatial dependencies while preserving positional information concurrently. The technology of Coordinate Attention entails concatenation observed by means of subsequent 1×1 convolution. The vertical and horizontal spatial records are encoded via BatchNorm and non-linear activation. This encoded data, segmented, undergoes every other 1×1 convolution to alter the channel measurement of the interest map to align with the input. The whole method concludes with normalization and weighted fusion by use of the sigmoid function, expressed as:

$$y_c(i, j) = x_c(i, j) \cdot g_c^h(i) \cdot g_c^w(j),$$

where $x_c(i, j)$ is the input feature map, and where $g_c^h(i)$ and $g_c^w(j)$ are the attention weights in the two spatial directions.

The integration of the Channel Attention module addresses the challenge of incomplete feature extraction arising from significant variations in forest fire images. The CA module enables the convolutional network to concentrate on learning crucial regions within forest fire elements, thereby comprehending the entire image more effectively by preserving long-range dependency information. The seamless fusion of the CA module with YOLOv8's backbone network successfully extends the receptive field, extracting more comprehensive positional information. This enhancement enables the model to adeptly adapt to multi-scale forest fire images.

2.3.4. WIoU Loss

In the context of woodland-hearth goal detection using unmanned-car aerial imagery, challenges arise, specifically in detecting small-sized fire site sources or smoke. YOLOv8 employs DFL and CIoU [31] for bounding box regression, yet CIoU has several drawbacks: it lacks balance between difficult and easy samples, inaccurately penalizes aspect ratios, and involves inverse trigonometric functions in computations. EIoU [32] improves upon CIoU by way of introducing separate penalties for width and height, more precisely reflecting variations between the floor reality and the anticipated boxes. SIoU [33] introduces an angle as a penalty factor, restricting regression freedom and enhancing model convergence speed.

WIoU represents a loss function for bounding box regression, incorporating a dynamic focus mechanism that is non-monotonic in nature. There are three versions of Wiou, among which v1 uses distance as an attention indicator, v2 reduces the weight of easy examples by constructing monotonic focus coefficients, and v3 considers the outlier values of anchor box quality, dynamically adjusting weights.

WIoU v1 improves mannequin generalization by decreasing the penalty on geometric measurements; v2 introduces a monotonic center of attention coefficients but reveals slower convergence. WIoU v3 defines an outlier price β , dynamically optimizing the weights for excessive and low-quality anchor bins through a non-monotonic focal point thing r , bettering common performance. In this calculation, the measurement of β is considered, meaning great anchor packing containers have a smaller weight in the loss, decreasing hazardous gradients for low-quality anchor boxes. The components for WIoU v1 are proven in Equations (1)–(3):

$$L_{WIoUv1} = R_{WIoU} \times L_{IoU}, \quad (1)$$

$$R_{WIoU} = \exp\left(\frac{(b_{cx} - b_{cx}^{gt})^2 + (b_{cy} - b_{cy}^{gt})^2}{c_w^2 + c_h^2}\right), \quad (2)$$

$$L_{IoU} = 1 - IoU. \quad (3)$$

The formula for WIoU v2 is given in Equation (4):

$$L_{WIoUv2} = \left(\frac{L_{IoU}^*}{L_{IoU}}\right)^\gamma \times L_{WIoUv1}, \gamma > 0. \quad (4)$$

The method for WIoU v3 is introduced in Equations (5)–(7). In Equation (6), the parameters δ and α are adjustable hyperparameters that can be tailored to exceptional models.

$$L_{WIoUv3} = r \times L_{WIoUv1}, \quad (5)$$

$$r = \frac{\beta}{\delta \alpha^{\beta - \delta}}, \quad (6)$$

$$\beta = \frac{L_{IoU}^*}{L_{IoU}} \in [0, +\infty). \quad (7)$$

Through comparison with various mainstream loss functions, WIoU v3 amalgamates the advantages of EIoU and SIoU. It incorporates a non-monotonic mechanism dynamically to account for anchor box quality, enabling the model to concentrate more on anchor boxes of standard quality, thereby improving target localization ability. In the context of forest fire target detection, WIoU v3 dynamically adjusts the loss weights for small targets, enhancing the model's performance.

3. Results

3.1. Training

The experimental hardware and software conditions are detailed in Table 1. Table 2 encompasses the parameters of the enhanced woodland-fire-classification detection model, which were initially configured through a preliminary exploration of the default settings of YOLOv8 and subsequently adjusted during the experimental process. Initially, the YOLOv8 model's learning rate was initialized to 0.01, and the training epochs were adaptively modified depending on the model's performance on the validation set, with 200 epochs determined as the optimal selection. The dataset for forest fire classification was partitioned into training, validation, and test sets in an 8:1:1 ratio.

Table 1. Experimental setting.

Device	Configuration
CPU	13th Gen Intel(R) Core(TM) i9-13900K
GPU	NVIDIA GeForce RTX 4090
System	Windows 10
Framework	Pytorch 2.0.0
IDE	Pycharm 2022.2.2
Python version	version 3.10.9

Table 2. Enhanced Model Training Parameter Setup for Forest Fire Detection.

Device	Configuration
Img-size	640 × 640
Epochs	200
Batch-size	8
Initial learning rate	0.01
Optimizer	SGD

3.2. Evaluation Metrics

The main research objective is to classify instances as either positive or negative, where fire and non-fire instances are categorized accordingly. In this context, there are four possible scenarios: True Positive, accurately predicting fires; True Negative, correctly predicting non-fires; False Positive, erroneously identifying non-fires as fires; and False Negative, misclassifying fires as non-fires.

Accuracy serves as a crucial indicator for assessing the model's predictive performance, quantified by calculating the ratio of True Positive and True Negative instances to the total expected positive and negative instances. The calculation method is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Recall, alternatively termed sensitivity or True Positive Rate, assesses the model's ability to identify positive instances compared to the actual positives. In forest fire detection, a high recall signifies the algorithm's proficiency in identifying fire occurrences accurately. The calculation method is as follows:

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

Precision, also known as Positive Predictive Value, measures the ratio of True Positive instances among those labeled as positive. In forest fire detection, high precision suggests

that the model can more accurately identify fire instances, reducing the probability of erroneously classifying non-fire instances as fires. The calculation approach is given by:

$$\text{Precision} = \frac{TP}{TP + FP}.$$

Assessing the model's efficacy in forest fire detection necessitates considering Recall and Precision to ensure efficient fire element detection and precise classification simultaneously.

Average Precision (*AP*) emerges as a crucial performance measure, providing a comprehensive evaluation of the Precision–Recall trade-off. The formula for *AP* calculation is as follows:

$$AP = \int_0^1 P(r) dr,$$

where the calculation of *AP* involves weighting and averaging the Average Precision values for different classes to obtain *mAP*:

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

where *n* represents the total number of categories, and where *AP_i* is the mean precision value for the *i*-th category. In this study, as forest fires belong to a single category, *AP* was employed as the overall performance assessment metric, with an IOU threshold set at 50%, denoted as *AP@50*.

Frames per Second (FPS) represents the rate at which the model processes images during target detection, providing a measure of the model's detection speed. The interplay between accuracy and FPS can be assessed by measuring the real-time performance of the model in detection tasks, with FPS and accuracy mutually influencing the model's feasibility in practical applications.

3.3. Experimental Comparison

This study introduces a bounding box regression loss function, WIoUv3, into the YOLOv8 forest fire detection model. Three alternative bounding box regression loss modules—GIoU, DIoU, and CIoU—are proposed and were compared with WIoU to assess their effectiveness. The objective was to determine the necessity of these enhancement modules and to understand their individual impacts on the performance of the fire detection model. Evaluation of the adequately trained models using the same dataset in our experiments and obtaining corresponding metrics yielded ablation study results as shown in Table 3. A thorough examination of the ablation study results provides insights into the impact of each enhancement module on model performance, offering valuable data for subsequent optimization and adjustments.

Table 3. The data of the ablation experiments.

Models	mAP@50	Precision	Recall	FPS	F1
YOLOv8	84.3	83.4	82.1	138	82.74
YOLOv8+GIoU	84.6	83.6	82.5	137	83.04
YOLOv8+DIoU	84.5	83.5	82.5	138	83.00
YOLOv8+WIoUv3	85.1	83.9	82.7	136	83.29
+DCN_C2f	87.3	87.4	82.7	115	84.99
+SCConv_C2f	87.3	87.4	82.8	137	85.03
+CA	90.2	90.1	83.5	135	86.67

In the first experiment, we combined YOLOv8 with WIoUv3 and further introduced the *DCN_C2f* module. By adjusting the C2f network structure and integrating the concept

of Deformable Convolution, the network was adapted to dynamically adjust its receptive field. The experimental results demonstrated a significant improvement in mAP@0.5, increasing from 84.3% to 87.3%. This indicates that the *DCN_C2f* module successfully encouraged the model to more effectively capture target features, especially in handling complex forest fire scenes. The introduction of *SCConv_C2f* in the second experiment did not further enhance performance, but such an improvement greatly increased the overall speed of the model.

In the third experiment, the implementation of the CA mechanism aimed to enhance the capture of directional and position-sensitive details regarding the targets. The experimental results demonstrated a notable enhancement in Precision and F1 Score, achieving 90.1% and 86.67%, respectively. Compared to the original model, Precision increased by 8.0%, while the F1 score increased by 4.7%. However, there was a slight decrease in FPS, indicating a slowdown in model processing speed. These findings suggest that CA effectively improves target detection, especially in forest fire scenarios, where directional and position-sensitive attributes are crucial for precise identification.

Despite the significant improvement in performance, the introduction of new modules usually comes with an increase in computational complexity. However, through *SCConv_C2f*, the model's FPS did not significantly decrease, indicating that our model maintains high performance while possessing a certain level of computational efficiency. These results indicate that the proposed improvement modules positively influence forest fire detection performance at different levels, providing a better solution for the practical application of the model.

To comprehensively evaluate the overall performance of the proposed enhanced YOLOv8 model in forest fire detection tasks, this study conducted comparative experiments with other mainstream object detection models, such as YOLOv3, YOLOv5, NAS-FPN, and FSAF. The comparative experimental results are illustrated in Figure 9. The experimental findings clearly demonstrate the significant superiority of the improved YOLOv8 model in terms of forest fire detection accuracy over other mainstream models, including the widely discussed YOLOv5 and YOLOv7. Additionally, the model's exceptionally high FPS values highlight its outstanding real-time performance. This overall performance enhancement establishes a solid foundation for the extensive application of the accelerated YOLOv8 model in practical scenarios, particularly in urgent situations requiring efficient and high-precision object detection. It is worth emphasizing that this performance advantage is not limited solely to forest fire detection tasks but also showcases the robust scalability of the model, offering broad possibilities for its application in other object detection tasks.

3.4. Visual Analysis

The evaluation of forest fire target detection outcomes between the enhanced YOLOv8 model and the unmodified YOLOv8 model showed a substantial improvement in the improved model's detection performance. Specifically, the augmented model demonstrated enhanced capabilities in resisting complex background interference and extracting global information related to forest fire targets. Notably, instances of missed detection and false alarms were significantly reduced, particularly showcasing optimal detection performance for small-scale forest fires. Moreover, the model exhibited commendable detection performance under varying lighting conditions, emphasizing its robustness to changes in illumination and further validating the effectiveness of its individual components. In summary, the proposed model proved to be more suitable for the task compared to the unmodified YOLOv8 model.

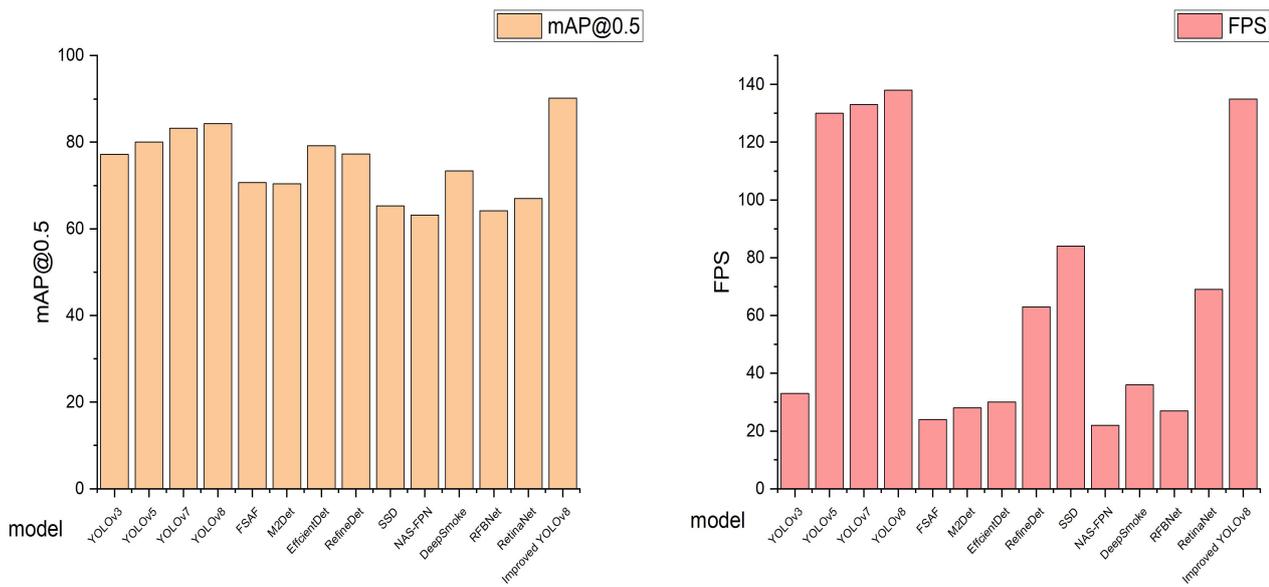


Figure 9. Improved YOLOv8 vs. the State-of-the-Art Methods: Bar Charts of FPS and mAP@0.5.

To better showcase the feasibility of the model, a selection of images from the test set was utilized for demonstration. The comparative detection results are illustrated in Figure 10, with the left image representing the unmodified model and the right image displaying our detection model. In scenarios where there was interference resembling forest fire targets, YOLOv8 incorrectly identified such targets as forest fire targets, a shortcoming addressed by the improved YOLOv8.



(a) Unmodified YOLOv8

(b) Modified YOLOv8

Figure 10. Model Feasibility Comparison—Enhanced YOLOv8 vs. Original Model in Challenging Fire-Like Object Scenarios.

Examining the detection results in Figure 11, the YOLOv8 model displayed misfit detection of rectangular box positions for forest fires. Particularly when faced with forest fire targets of varying scales, YOLOv8 exhibited multiple instances of missed detection, leading to suboptimal performance. In contrast, the improved YOLOv8 model aligned more accurately with the actual framework of forest fire targets, eliminating instances of missed detection and improving detection performance for forest fire types.

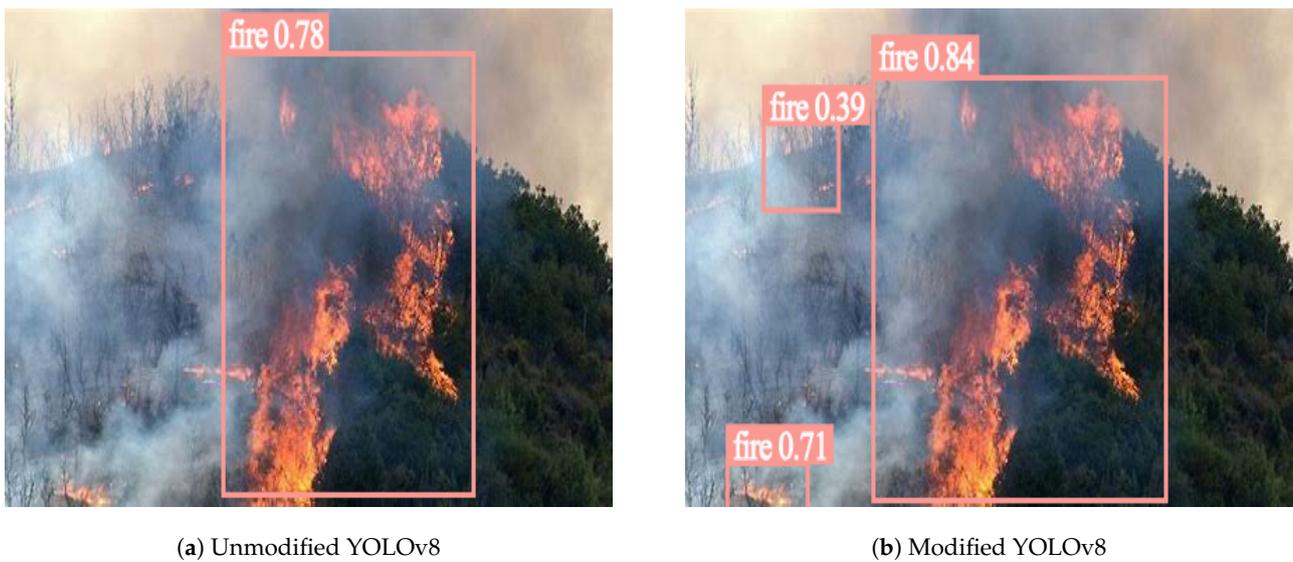


Figure 11. Comparative Detection Results—Performance of YOLOv8 and Enhanced YOLOv8 in Detecting Forest Fire Targets at Different Scales.

Promptly identifying initial small-scale forest fires holds paramount importance in forest fire detection. As depicted in Figure 12, YOLOv8 encountered difficulties in discerning small-scale forest fire targets at a distance. In contrast, the improved YOLOv8 showcased precise detection in the image, attaining a confidence level of 0.92.



Figure 12. Comparison of Early Small Target Forest Fire Detection: YOLOv8 vs. Modified YOLOv8.

As shown in Figure 13, YOLOv8 encountered difficulties in adequately capturing information about forest fire targets in the image. This challenge led to bounding boxes being restricted to localized areas of the forest fire targets, resulting in inadequate positioning. Conversely, the enhanced YOLOv8 effectively extracted vital details regarding forest fire targets, presenting a more comprehensive perceptual scene.

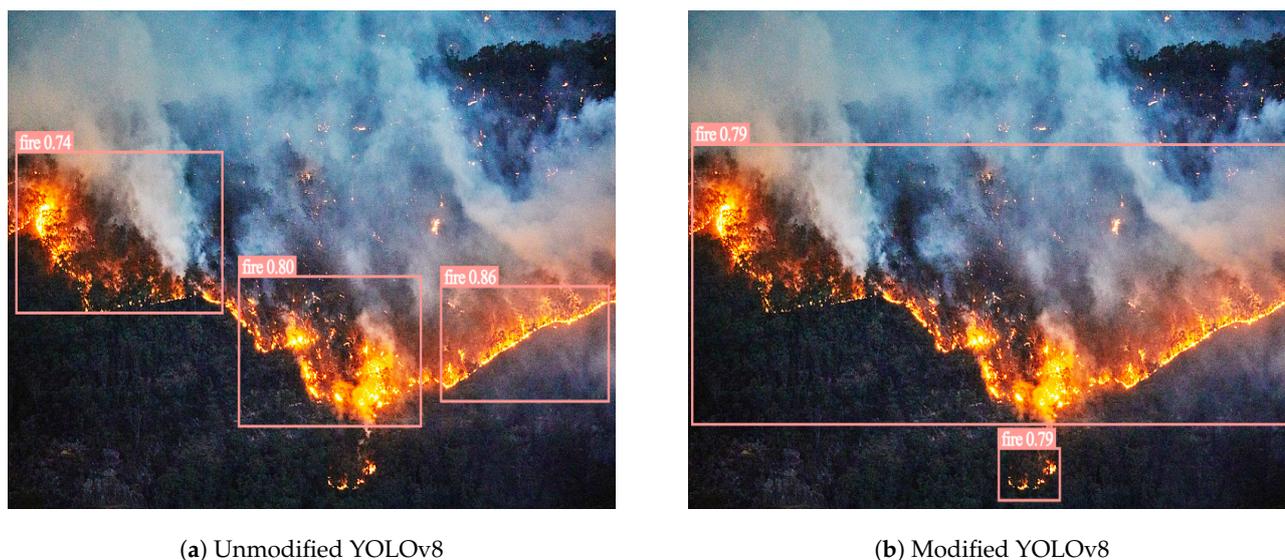


Figure 13. Comparative Global Information Extraction—Performance of YOLOv8 vs. Improved YOLOv8 in Forest Fire Target Detection.

4. Discussion

The importance of forests to human society cannot be overlooked; hence, the well-timed detection and mitigation of wooded-area fires to minimize their effect on the environment are especially urgent. However, in contrast to other frequent objects, flames showcase a higher morphology and complicated dynamic characteristics. Additionally, the mutual occlusion of timber throughout the whole wooded area makes taking pictures of these points significantly challenging. The current detection that applied sciences go through suffers from a number of deficiencies, making it tough to successfully tackle these issues. Traditional techniques for detecting wooded-area fires require the use of manually created recognizers. These strategies fail to extract the imperative points of flames, resulting in overall subpar detection performance and slow detection speeds.

In deep learning-based detection approaches, the two-stage goal-detection models, such as Faster R-CNN [34], demand prolonged training periods and detection times, often failing to meet real-time detection requirements. While single-stage goal detection keeps real-time demands, precision barely decreases. SSD [35] and the YOLO series, exemplified with the aid of the complexity of the SSD debugging system, rely closely on guide expertise. In comparison, the YOLO series, especially YOLOv8, stands out in wooded-area hearth-detection strategies, due to its benefits, such as a compact mannequin size, low deployment costs, and quick detection speed. However, due to massive variants in the scale of woodland fires, obtaining satisfactory cognizance of the consequences remains difficult in the case of multi-scale wooded-area fire-site images. Therefore, wooded-area furnace detection remains a difficult look-up area.

To tackle these challenges, this paper proposes an algorithmic mannequin primarily based on the improved YOLOv8 by adjusting the C2f community shape and integrating it one-by-one with Deformable Convolution and SCConv to shape *DCN_C2f* and *SCConv_C2f*. This approach involves augmenting the Channel Attention interest mechanism, modifying backbone network modules, improving the loss function, and employing various techniques. Our experimental results demonstrate that the proposed model exhibits high average precision and rapid frames per second, rendering it applicable to forest fire detection across various types and scales.

The improved YOLOv8 demonstrates high accuracy in forest fire detection; however, there is still room for improvement. Primarily, while fusion modules within the network contribute to enhanced precision, they may inadvertently compromise detection speed. Furthermore, despite comprising 2692 images depicting various scenes and flame types,

our research dataset remains relatively small in scale. The operational capabilities and the field of view of the drones may also be limited, especially in complex terrains or harsh weather conditions, such as thick fog or strong winds. This could potentially impact the accuracy and efficiency of fire detection. In our future research endeavors, we will focus on optimizing the model and deploying it on unmanned aerial vehicles equipped with different cameras. Simultaneously, we plan to optimize the dataset to enhance the accuracy of fire detection. Chen et al. [36] proposed a method utilizing a multi-modal dataset collected by drones, achieving high accuracy in detecting fire and smoke pixels by collecting dual-channel videos containing RGB and thermal images. Furthermore, they provided rich auxiliary data, such as georeferenced point clouds, orthomosaics, and weather information to offer more comprehensive contextual information. This served as inspiration for our work. Mashraqi Aisha M. et al. [37] emphasized the design of forest fire detection and classification in drone imagery, using the modified deep learning (DIFFDC-MDL) model. Remarkably, they employed the Shuffled Frog Leaping algorithm and simulated the results using a database containing fire and non-fire samples, thereby enhancing the classification accuracy of the DIFFDC-MDL system. In the future, we will explore the use of metaheuristic algorithms to further improve the model.

5. Conclusions

Global forest fires are increasingly frequent, and the challenges associated with their control have risen. Failure to address them promptly results in significant financial losses and safety issues. Flames exhibit a greater diversity and complexity in both morphology and dynamic features compared to other objects, and this is particularly challenging, due to the mutual occlusion of trees in forests. Hence, the identification and recognition of fire types, coupled with corresponding fire suppression measures, hold not only developmental potential but also practical importance.

In this work, an enhanced model for forest fire classification and detection is presented, leveraging the YOLOv8 architecture. By modifying the C2f network structure and integrating Deformable Convolution and SCConv, resulting in *DCN_C2f* and *SCConv_C2f*, the model achieves adaptive adjustment of the receptive field. This leads to a reduction in parameters and FLOPs while enhancing feature representation. The incorporation of Coordinate Attention enhances the Detect section, improving the capturing and utilization of features, consequently enhancing fire detection accuracy. WIoU v3 is incorporated into the loss function for bounding box regression, utilizing a dynamic mechanism that is non-monotonic to establish a more rational distribution strategy for gradient gain. Our experimental results showcased the model's superiority in forest fire performance compared to the YOLOv8 algorithm. Our future research will focus on optimizing the model, particularly in balancing accuracy and detection speed within the fusion module, and deploying the model on drones equipped with different cameras. Additionally, our plans involve expanding the dataset size to enhance fire detection accuracy.

Author Contributions: Conceptualization, Z.Y., Y.S., Y.W. and J.L.; Data Curation, Z.Y.; Formal Analysis, Z.Y. and Y.S.; Funding Acquisition, J.L.; Investigation, Z.Y.; Methodology, Z.Y. and Y.S.; Project Management, Y.S.; Resources, Y.S.; Software, Z.Y.; Supervision, J.L.; Validation, Y.W.; Visualization, Y.S.; Writing—Original Draft, Z.Y.; Writing—Review and Editing, Z.Y. and Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, Grant No. 12361029, and the Chinese National College Students Innovation and Entrepreneurship Training Program, Grant No. DC20231669.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the conclusions of this study are accessible from the corresponding author upon reasonable request. The data are not publicly available due to privacy.

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

References

- Führer, E. Forest functions, ecosystem stability and management. *For. Ecol. Manag.* **2000**, *132*, 29–38. [CrossRef]
- Li, Q. Effect of forest bathing trips on human immune function. *Environ. Health Prev. Med.* **2010**, *15*, 9–17. [CrossRef]
- Thonicke, K.; Venevsky, S.; Sitch, S.; Cramer, W. The role of fire disturbance for global vegetation dynamics: Coupling fire into a Dynamic Global Vegetation Model. *Glob. Ecol. Biogeogr.* **2001**, *10*, 661–677. [CrossRef]
- Yadav, V.S.; Yadav, S.S.; Gupta, S.R.; Meena, R.S.; Lal, R.; Sheoran, N.S.; Jhariya, M.K. Carbon sequestration potential and CO₂ fluxes in a tropical forest ecosystem. *Ecol. Eng.* **2022**, *176*, 106541. [CrossRef]
- Watson, A.J.; Lovelock, J.E. The dependence of flame spread and probability of ignition on atmospheric oxygen: An experimental investigation. *Fire Phenom. Earth Syst. Interdiscip. Guide Fire Sci.* **2013**, 273–287. [CrossRef]
- Bouabdellah, K.; Noureddine, H.; Larbi, S. Using wireless sensor networks for reliable forest fires detection. *Procedia Comput. Sci.* **2013**, *19*, 794–801. [CrossRef]
- Yuan, C.; Zhang, Y.; Liu, Z. A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques. *Can. J. For. Res.* **2015**, *45*, 783–792. [CrossRef]
- Chuvieco, E.; Aguado, I.; Salas, J.; García, M.; Yebra, M.; Oliva, P. Satellite remote sensing contributions to wildland fire science and management. *Curr. For. Rep.* **2020**, *6*, 81–96. [CrossRef]
- Chen, Y.; Zhang, Y.; Xin, J.; Wang, G.; Mu, L.; Yi, Y.; Liu, H.; Liu, D. UAV image-based forest fire detection approach using convolutional neural network. In Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 19–21 June 2019; pp. 2118–2123.
- Mahmoud, M.A.I.; Ren, H. Forest fire detection and identification using image processing and SVM. *J. Inf. Process. Syst.* **2019**, *15*, 159–168.
- Emmy Prema, C.; Vinsley, S.; Suresh, S. Efficient flame detection based on static and dynamic texture analysis in forest fire detection. *Fire Technol.* **2018**, *54*, 255–288. [CrossRef]
- Han, X.F.; Jin, J.S.; Wang, M.J.; Jiang, W.; Gao, L.; Xiao, L.P. Video fire detection based on Gaussian Mixture Model and multi-color features. *Signal Image Video Process.* **2017**, *11*, 1419–1425. [CrossRef]
- Mohajane, M.; Costache, R.; Karimi, F.; Pham, Q.B.; Essahlaoui, A.; Nguyen, H.; Laneve, G.; Oudija, F. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecol. Indic.* **2021**, *129*, 107869. [CrossRef]
- Mahaveerakannan, R.; Anitha, C.; Thomas, A.K.; Rajan, S.; Muthukumar, T.; Rajulu, G.G. An IoT based forest fire detection system using integration of cat swarm with LSTM model. *Comput. Commun.* **2023**, *211*, 37–45. [CrossRef]
- Barmpoutis, P.; Dimitropoulos, K.; Kaza, K.; Grammalidis, N. Fire detection from images using faster R-CNN and multidimensional texture analysis. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 8301–8305.
- Alice, K.; Thillaivanan, A.; Rao, G.R.K.; Rajalakshmi, S.; Singh, K.; Rastogi, R. Automated Forest Fire Detection using Atom Search Optimizer with Deep Transfer Learning Model. In Proceedings of the 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 4–6 May 2023; pp. 222–227.
- Lin, J.; Lin, H.; Wang, F. A semi-supervised method for real-time forest fire detection algorithm based on adaptively spatial feature fusion. *Forests* **2023**, *14*, 361. [CrossRef]
- Ghali, R.; Akhloufi, M.A.; Mseddi, W.S. Deep learning and transformer approaches for UAV-based wildfire detection and segmentation. *Sensors* **2022**, *22*, 1977. [CrossRef]
- Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 764–773.
- Li, J.; Wen, Y.; He, L. SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 6153–6162.
- Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 13713–13722.
- Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051.
- Chino, D.Y.; Avalhais, L.P.; Rodrigues, J.F.; Traina, A.J. Bowfire: Detection of fire in still images by integrating pixel color and texture analysis. In Proceedings of the 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images, Salvador, Brazil, 26–29 August 2015; pp. 95–102.
- Ultralytics. YOLOv8. Available online: <https://docs.ultralytics.com/> (accessed on 21 June 2023).
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
- JOCHER. Network Data. 2020. Available online: <https://github.com/ultralytics/yolov5> (accessed on 24 December 2022).
- Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
- Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.

29. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. mixup: Beyond empirical risk minimization. *arXiv* **2017**, arXiv:1710.09412.
30. Shen, L.; Lang, B.; Song, Z. DS-YOLOv8-Based Object Detection Method for Remote Sensing Images. *IEEE Access* **2023**, *11*, 125122–125137. [[CrossRef](#)]
31. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 12993–13000.
32. Zhang, Y.F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [[CrossRef](#)]
33. Gevorgyan, Z. SloU Loss: More Powerful Learning for Bounding Box Regression. *arXiv* **2022**, arXiv:2205.12740.
34. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer, Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
35. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14; pp. 21–37.
36. Chen, X.; Hopkins, B.; Wang, H.; O’Neill, L.; Afghah, F.; Razi, A.; Fulé, P.; Coen, J.; Rowell, E.; Watts, A. Wildland Fire Detection and Monitoring Using a Drone-Collected RGB/IR Image Dataset. *IEEE Access* **2022**, *10*, 121301–121317. [[CrossRef](#)]
37. Mashraqi, A.M.; Asiri, Y.; Algarni, A.D.; Abu-Zinadah, H. Drone imagery forest fire detection and classification using modified deep learning model. *Therm. Sci.* **2022**, *26*, 411–423. [[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.