IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication February 25, 2025, date of current version December 29, 2024.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Fire and Smoke Detection Based on Improved YOLOV11

ZHIPENG XUE[1], LINGYUN KONG[2], HAIYANG WU[3], JIALE CHEN[4]
[1]Xijing University, Xi'an, Shaanxi 710123, China (e-mail: 3213924681@qq.com)
[2]Xijing University, Xi'an, Shaanxi 710123, China (e-mail: 1400100383@qq.com)
[3]Xijing University, Xi'an, Shaanxi 710123, China(e-mail: 2010530510@qq.com)
[4]Xijing University, Xi'an, Shaanxi 710123, China(e-mail: 1350171520@qq.com)

Corresponding author: Lingyun Kong (e-mail: 1400100383@qq.com).

ABSTRACT Fire and smoke detection is an important measure to ensure the safety of people's lives and property, as well as a crucial link in maintaining ecological balance and supporting scientific research. Traditional object detection methods rely more on manually designed features and rules. Although they are relatively simple to implement, their performance is limited in complex and variable practical applications. In contrast, deep learning-based methods can automatically learn deep features in data and have higher accuracy and stronger generalization ability. However, complex backgrounds, large environmental changes, and data requirements pose great challenges to high-precision outdoor smoke detection. To address these issues, this paper proposes an improved model, YOLOV11-DH3, based on YOLOV11. In this paper, the core DCN2 (Deformable Convolutional Networks2) of the YOLOV11 Head is replaced with the DCN3 module to form a new detection head. In addition, the loss function CIOU in YOLOV11 is replaced with IOU to consider the irregular shape of fire and smoke and the problem of multi-scale targets. To evaluate the performance of the algorithm, comprehensive experiments were conducted on two distinct datasets: a public fire and smoke dataset provided by Baidu Paddle featuring close-range views and a public wildfire smoke dataset from the YOLO official website with distant outdoor perspectives. The experimental results show that on the Baidu Paddle dataset, the average accuracy of the model is improved by 1.4% compared to the original model, reaching 58%, the F1 score is improved by 2%, reaching 58%, with a precision of 91.6% and recall of 90%. Our cross-dataset analysis provides valuable insights into model performance across different detection scenarios. The proposed model demonstrates the ability to accurately detect fire and smoke in complex backgrounds, and this progress is of great significance for protecting people's lives and maintaining ecological balance.

INDEX TERMS Object detection; YOLOv11-DH3; IOU; Fire and smoke detection; Deep learning

I. INTRODUCTION

In modern society, fireworks are widely used in celebrations and public festivals. However, the use of fireworks also brings potential safety hazards, especially in densely populated areas and thick forest environments. Therefore, fire prevention, emergency response, and safety monitoring play a crucial role. Fires not only pose a serious threat to human life and property safety but also may have a long-term impact on the environment and ecosystem [24]. Therefore, timely and accurate detection of the early signs of a fire is the key to effectively preventing the spread of the fire and reducing losses.

Statistics show that the Australian bushfires from 2019 to 2020 burned 100 million hectares of forest, resulting in the death of millions of animals and an economic loss of more than 4 billion dollars [1]; on November 15, 2010, a particularly serious fire accident occurred in an apartment building at No. 728 Jiaozhou Road, Jing'an District, Shanghai, killing 58 people and injuring 71 people, with a direct economic loss of 158 million yuan [2]; on June 3, 2013, a particularly serious fire and explosion accident occurred in the main workshop of Jilin Baoyuanfeng Poultry Industry Co., Ltd. in Dehui City, Jilin Province, killing 121 people and injuring 76 people, with a direct economic loss of 182 million yuan [3]. These examples demonstrate the huge losses caused by fires, involving not only property and life losses but also environmental, social, and economic impacts. Factors such as the spread speed of the fire, the fire resistance of buildings, and the efficiency of emergency response all determine the final scale and impact of the disaster. Therefore, it is increas-

ingly important to improve the ability of fire monitoring, early warning, and response, especially in complex and dangerous environments.

In fire monitoring and prevention, the detection of smoke and flame is a crucial step. The core task of smoke and flame detection is to detect fires early [17], issue alarms in a timely manner, and provide accurate information for subsequent fire extinguishing and evacuation, so as to maximize the protection of personnel safety, reduce property losses, and provide effective support for rescue operations [12].

Traditional smoke and flame detection methods mainly rely on physical and chemical principles to identify the signs of a fire [18]. They are easily affected by non-fire factors such as steam, dust, and cooking oil fumes, resulting in false alarms. Moreover, traditional detectors usually cannot identify the characteristics of smoke or flame produced by different substances burning, reducing the pertinence of early warnings. In addition, some types of traditional detectors (such as ionization smoke detectors) contain radioactive materials and require special treatment, and all detectors need to be regularly cleaned and calibrated to ensure normal operation. Therefore, the research on smoke and flame object detection methods has practical application value and significance, enabling people to more accurately distinguish between real fires and non-fire situations [25].

With the continuous progress of computer vision technology, it has been widely used in fire and smoke detection. Based on computer vision technology, researchers have conducted in-depth research on fire and smoke object detection and proposed various solutions, including traditional digital image processing and deep learning methods [26].

II. RELATED WORK

A. FLAME DETECTION BASED ON DIGITAL IMAGE PROCESSING

1) Flame Detection Based on Color Features

Flame detection methods based on color features [4] usually identify flames by analyzing the color information in images or videos. Flames usually have distinct color features such as orange, red, yellow, and blue, which are significantly different from the color features of other objects.

2) Flame Detection Based on Texture Features

Flame detection methods based on texture features [10] identify flames by analyzing the texture characteristics of the flame region. The texture features of flames are significantly different from those of the surrounding background objects. The main methods include Local Binary Pattern (LBP). The texture of flames usually appears as irregular spots and stripes, and LBP can effectively extract these features and distinguish flames from other objects; Gray-Level Co-occurrence Matrix (GLCM) [20]. The GLCM of flames usually shows low contrast and high entropy, which is different from the texture features of background objects. By calculating the statistical features of GLCM (such as contrast, correlation, energy, etc.), a texture model of flames can be constructed.

3) Flame Detection Based on Motion Information

Flame detection methods based on motion information identify flames by analyzing the dynamic characteristics of flames. Flames are usually dynamic, with rapidly changing shapes and positions, while static background objects are relatively stable. By analyzing the motion information in the image sequence, flames can be effectively distinguished from other objects. Specific methods include optical flow method, which is a method for estimating the pixel motion in an image sequence; frame difference method, which detects moving objects by comparing the differences between adjacent frames.

4) Smoke and Flame Detection Based on Shape Features

Smoke and flame detection methods based on shape features identify flames and smoke by analyzing their geometric shapes. The shape features of flames and smoke are significantly different from those of other objects, especially in the early stage of a fire, the shape features of smoke and flame are more obvious. Specific methods include edge detection algorithms (such as Canny edge detection, Sobel operator, etc.) to extract the contours of flames and smoke; morphological operations to remove noise and enhance the shape features of flames and smoke, thereby improving the detection accuracy.

Although traditional flame and smoke detection methods based on color, texture, motion, and shape features have shown certain effectiveness in some scenarios, their robustness and accuracy in complex environments are often inferior to those of deep learning-based methods [27].

B. FIRE AND SMOKE DETECTION METHODS BASED ON DEEP LEARNING

1) Convolutional Neural Network (CNN)

The convolutional neural network method [5] includes one-stage detectors and two-stage detectors. One-stage detectors such as the YOLO series directly predict the object bounding boxes and categories in an image through a single neural network, with the characteristics of fast speed and high accuracy. Two-stage detectors such as Faster R-CNN generate candidate regions through a Region Proposal Network (RPN) and then classify and regress these regions [28].

2) Spatio-Temporal Convolutional Network (3D-CNN and 2+1D-CNN)

3D-CNN extends the convolution operation in the time dimension and can capture the spatio-temporal information in video sequences [6]. Flames and smoke usually change dynamically, and 3D-CNN can effectively utilize this characteristic to improve the detection accuracy; 2+1D-CNN is a method that combines two-dimensional spatial convolution and one-dimensional time convolution, which can extract spatio-temporal features while maintaining high computational efficiency.

3) Generative Adversarial Network (GAN)

CycleGAN can be used to generate realistic flame and smoke images [7] to expand the training dataset. By generating more diverse samples; CGAN can generate specific types of flame and smoke images according to given conditions (such as environment, weather, etc.), helping the model better adapt to different application scenarios. In addition, CGAN can also be used for anomaly detection by generating images of normal scenes to identify flames or smoke that do not conform to the normal scene.

4) Recurrent Neural Network (RNN) and Its Variants

LSTM [8] and GRU [9] are two commonly used recurrent neural networks that can handle the long-term dependencies in sequence data. In fire and smoke detection, they can be used to model the time evolution process of flames and smoke, especially in early fire warning, they can capture the weak changes of flames.

5) Fire and Smoke Detection Method Based on Transformer

Vision Transformer is a visual model based on the Transformer [11] architecture that can directly process image patches as input and capture the global dependencies in the image through the self-attention mechanism. ViT can capture the global dependencies between different regions in the image through the self-attention mechanism, which is very useful for detecting flames and smoke with irregular shapes and being easily affected by the environment. And by introducing a multi-scale feature fusion mechanism, ViT can extract the features of flames and smoke at different scales and improve the detection accuracy.

6) Lightweight Models for Specialized Object Detection

Recent advancements in deep learning have led to the development of specialized lightweight models designed for specific object detection tasks. Chen et al. [21] proposed EFS-YOLO, a lightweight network based on improved YOLOv8s architecture for steel strip surface defect detection. By employing EfficientViT as the backbone network and designing a lightweight C2f-Faster-EffectiveSE Block (CFE-Block), they effectively reduced model parameters by 49.5% and calculations by 62.7% while maintaining detection accuracy. Their approach demonstrates that domain-specific optimization can significantly enhance performance in industrial applications. Similarly, Hao et al. [22] introduced YOLO-CXR, a novel detection network specifically designed for locating multiple small lesions in chest X-ray images. Their network enhances the YOLOv8s backbone by replacing ordinary convolutional layers with RefConv layers and introducing an Efficient Channel and Local Attention (ECLA) mechanism. Additionally, they incorporated a dedicated small-lesion detection head and Selective Feature Fusion (SFF) technique to significantly improve the detection of lesions at different scales. These specialized approaches demonstrate the effectiveness of adapting and optimizing deep learning architectures for domain-specific object detection tasks with irregular shapes and varying scales, which is particularly relevant to fire and smoke detection challenges [29].

III. DATASET INTRODUCTION

In this research, we used two distinct datasets to thoroughly evaluate our proposed method:

A. BAIDU PADDLE DATASET

In this experiment, the public fire and smoke detection dataset provided by Baidu Paddle is used. This dataset contains 2059 fire and smoke images and includes the annotation information of each image. We divided the dataset into a training set of 1759 images and a validation set of 300 images for the experiment. This dataset features close-range views of various fire scenarios, including forest fires, building fires, and grassland fires, with visible flames and smoke of diverse shapes and sizes.

B. YOLO MONITORING DATASET

To further validate our approach in different environmental conditions, we incorporated another public dataset from the YOLO official website, which we refer to as the "Monitoring Perspective Dataset." This dataset contains 1846 images primarily captured from monitoring cameras positioned at distances from potential fire sources. The images feature early-stage smoke detection scenarios in mountainous and hill regions, with smoke regions typically occupying only about 5.7 percent of the total image area.

This dataset provides several unique characteristics compared to the Baidu Paddle dataset: 1. Long-distance monitoring perspectives versus close-range direct views 2. Primarily early-stage smoke detection versus various fire stages 3. Relatively regular smoke shapes versus highly irregular fire and smoke patterns 4. Less complex backgrounds (mainly sky and mountains) versus diverse environmental contexts

We similarly divided this dataset into a training set of 516 images and a validation set of 147 images. The inclusion of this second dataset allows us to comprehensively evaluate our model's performance across varying detection scenarios and spatial scales.

IV. ALGORITHM DESIGN

A. YOLO ARCHITECTURE EVOLUTION AND CORE PRINCIPLES

YOLO (You Only Look Once) is a family of real-time object detection algorithms first introduced by Redmon and Farhadi in 2015 [13]. Unlike two-stage detectors such as R-CNN variants that first propose regions and then classify them, YOLO adopts a unified, single-pass approach that directly predicts bounding boxes and class probabilities from full images in one evaluation. This design philosophy enables YOLO to achieve remarkable inference speed while maintaining competitive accuracy.

The core principle of YOLO lies in its grid-based prediction mechanism. The input image is divided into an $S \times S$ grid, where each grid cell is responsible for predicting objects

whose centers fall within it. Each cell predicts B bounding boxes, each box comprising five parameters: center coordinates (x, y), width (w), height (h), and confidence score. Simultaneously, the cell predicts C conditional class probabilities. This approach formulates object detection as a regression problem rather than a classification problem over sliding windows.

Through successive iterations (YOLOv2 through YOLOv11), the architecture has evolved significantly [30] [31]:

- **Backbone Evolution:** From the initial custom architecture in YOLOv1 to DarkNet variants, and in more recent versions, CSP (Cross-Stage Partial) networks that optimize gradient flow and reduce computational burden.
- **Feature Pyramid Networks (FPN):** Introduced to enhance multi-scale detection capabilities, allowing the model to detect objects of varying sizes effectively.
- **Anchor-based to Anchor-free:** Later versions like YOLOv8 moved towards anchor-free designs that simplify the prediction mechanism and reduce hyperparameter dependence.
- **Advanced Loss Functions:** Progression from simple mean squared error to IoU-based losses (GIoU, DIoU, CIoU) that better align with the evaluation metrics [33] [34].
- **Attention Mechanisms:** Integration of spatial and channel attention to enhance feature representation.

YOLOv11, the latest iteration developed by Ultralytics [19], represents a culmination of these evolutionary improvements. It introduces several architectural innovations:

1) **Enhanced Backbone:** YOLOv11 adopts an improved backbone with C3K2 modules replacing the CF2 modules used in YOLOv8, enhancing feature extraction while maintaining computational efficiency.
2) **C2PSA Module:** A key innovation in YOLOv11 is the introduction of the C2PSA module after the SPPF (Spatial Pyramid Pooling - Fast) layer. This module incorporates Pointwise Spatial Attention to enhance the model's ability to focus on relevant spatial features, particularly beneficial for detecting objects with irregular shapes like smoke and fire.
3) **Optimized Detection Head:** YOLOv11 incorporates ideas from YOLOv10's head design but implements them using depthwise separable convolutions to reduce computational overhead. The head structure employs DCNv2 (Deformable Convolutional Networks version 2) to adaptively adjust sampling locations based on input features.
4) **CIOU Loss Function:** YOLOv11 utilizes Complete Intersection over Union (CIOU) loss by default, which considers not only the overlap area but also the distance between centers and aspect ratio similarity between prediction and ground truth boxes.

Figure 1 illustrates the comprehensive network architecture of YOLOv11, highlighting the connections between different components and the information flow from input to output. The architecture consists of three main parts: the backbone for feature extraction, the neck for feature fusion across different scales, and the detection heads for final prediction. This design enables YOLOv11 to achieve state-of-the-art performance in object detection tasks while maintaining real-time inference speed, making it suitable for applications requiring both high accuracy and low latency [40].

In its detection head, the head idea of YOLOv10 is introduced into the head of YOLO11, and the depthwise separable method is used to reduce redundant calculations and improve efficiency.

B. LIMITATIONS OF YOLOV11

Firstly, since the core DCNV2 (Deformable Convolutional Networks) of the dynamic detection head of YOLOV11 introduces an additional offset learning mechanism, it needs to store additional offset parameters, increasing the amount of operation parameters and making it difficult to deploy. Secondly, although DCNv2 performs well in handling geometric changes, its detection performance for very small targets may not be as expected, and using DCNv2 may make the post-processing steps more complicated. For example, in an object detection task, a more complex non-maximum suppression (NMS) strategy may be required to handle redundant prediction boxes, especially in a scene with densely arranged small targets. In addition, the default loss function CIOU (Complete Intersection over Union) of YOLOV11 mainly focuses on the overall intersection over union (IoU) of two bounding boxes, including the outer region. This design may lead to the neglect of the overlapping area inside the target in some cases, and for small targets, the IoU calculation of CIOU may result in inaccurate positioning due to the relatively small size of the bounding boxes. To solve these problems, specific improvement methods are proposed in this paper, such as C.

C. YOLOV11-DH3

This paper proposes to upgrade and replace the core DCNv2 of the detection head of YOLOV11 to build a more lightweight object detection model. And replace the loss function to enhance its higher performance and better generalization ability in object detection tasks. The specific details are as follows:

1) Replace DCNV2 with DCNV3

Convolutional neural networks have achieved remarkable success in visual recognition tasks. Nevertheless, they still have shortcomings. A key challenge in visual recognition is how to adapt to geometric changes in object scale, pose, viewpoint, and part deformation or model geometric transformation. Fixed-size convolution kernel operations cannot achieve this effect well.To address these challenges, Jifeng Dai, Haozhi Qi, et al [14]. proposed deformable convolution. This innovative approach introduces learnable offsets to adjust the positions of the convolution kernels, enabling the model to better adapt to geometric variations in objects. As a result, deformable convolution significantly enhances

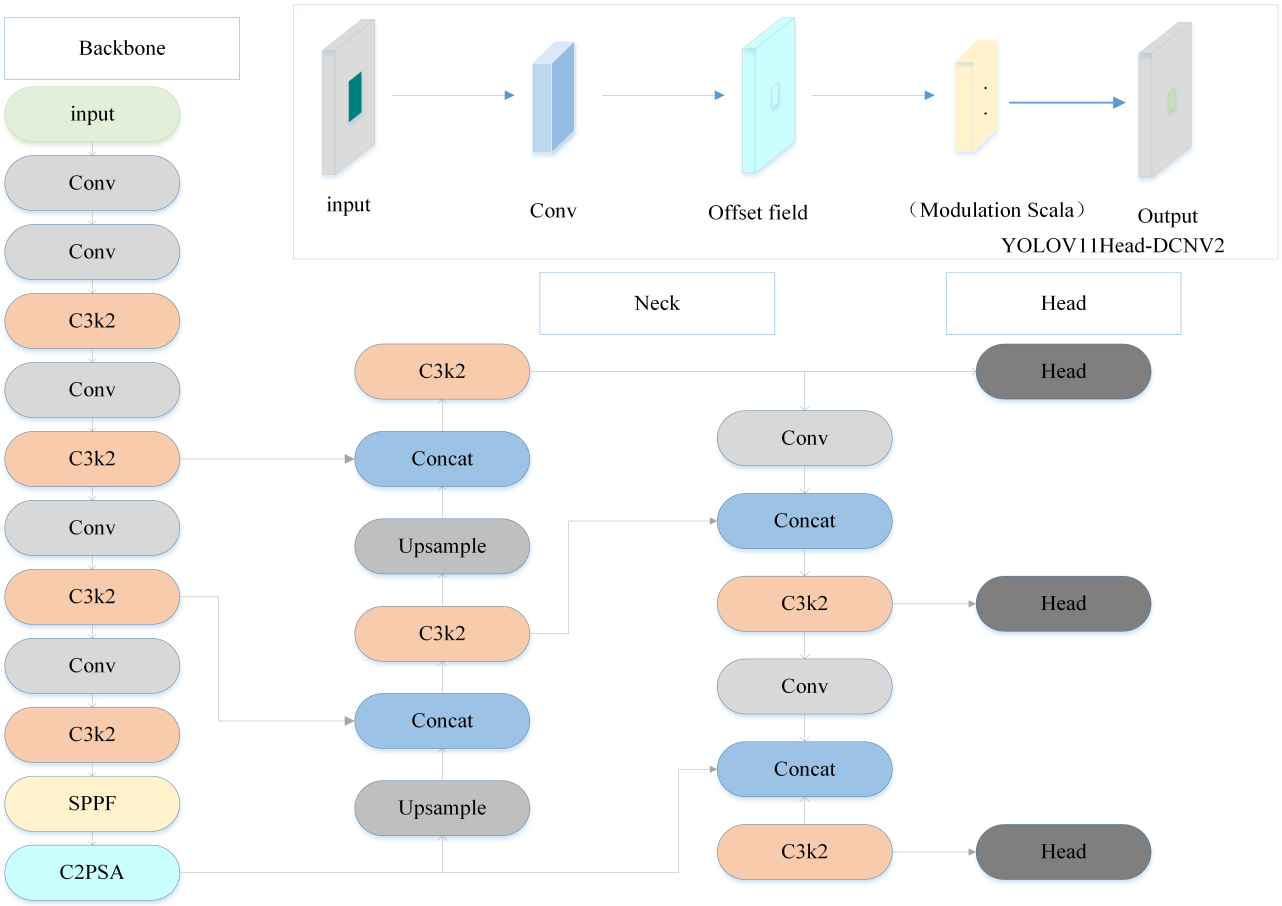

FIGURE 1. **Network structure diagram of YOLOv11.**

the performance of detection and recognition tasks. When deformable convolutions are stacked, the effect of composite deformation is profound. Its implementation process is shown in Figure 2. The left side is the input feature, and the right side is the output feature. Our convolution kernel size is 3x3. We map the 3x3 region in the input feature to 1x1 in the output feature. The problem is how to select this 3x3 region. Traditional convolution has a regular shape, while deformable convolution adds an offset and then calculates for each point separately, changing the selection of each point in the 3x3 region and extracting some points that may have richer features, thereby improving the detection effect [32].

a: Detailed Comparison between DCNv2 and DCNv3

The core distinction between our proposed YOLOv11-DH3 and the original YOLOv11 lies in the replacement of DCNv2 with DCNv3 in the detection head. To better understand the significance of this modification, we provide a detailed technical comparison of these two modules.

DCNv2, introduced by Dai et al. [14], enhances the standard convolution operation by allowing the sampling grid to be deformed adaptively based on input features. It achieves this through two key components, as shown in Equation (1):

$$\mathbf{y}(p) = \sum_{k=1}^{K} w_k \cdot x(p + p_k + \Delta p_k) \cdot \Delta m_k \tag{1}$$

where Δp_k represents the learned offset for the k-th sampling point, and Δm_k is the modulation scalar that adjusts the contribution of each sampling point. This mechanism enables DCNv2 to adapt to geometric variations in objects, making it particularly suitable for detecting objects with variable shapes.

DCNv3, proposed by Li et al. [16], extends this concept with several critical enhancements, as formulated in Equation (2):

$$\mathbf{y}(p_0) = \sum_{g=1}^{G} \sum_{k=1}^{K} \mathbf{w}_g \mathbf{m}_{gk} \mathbf{x}_g(p_0 + p_k + \Delta p_{gk}) \tag{2}$$

The key improvements in DCNv3 include:

- **Group-wise Deformation:** DCNv3 introduces a group-wise deformation mechanism (denoted by index g in Equation (2)), allowing different channel groups to have independent deformation patterns. This significantly enhances the model's ability to capture complex geometric transformations.

- **Enhanced Offset Prediction:** DCNv3 employs a more sophisticated offset prediction network that leverages multi-scale context information, resulting in more precise sampling location adjustments.
- **Dynamic Kernel Selection:** Unlike DCNv2, which uses a fixed kernel shape after deformation, DCNv3 dynamically adjusts the kernel shape based on input features, providing greater adaptability to varying object morphologies.
- **Improved Feature Integration:** DCNv3 incorporates advanced feature integration techniques that better preserve spatial relationships while allowing for flexible deformation.

Figure 2 illustrates the implementation process of both DCNv2 and DCNv3, showing how the enhanced deformation capabilities of DCNv3 enable more adaptive feature extraction. Specifically, for fire and smoke detection tasks, DCNv3's ability to dynamically adjust sampling locations with greater flexibility allows it to better capture the irregular boundaries and amorphous shapes characteristic of smoke and flames, especially in complex scene conditions.

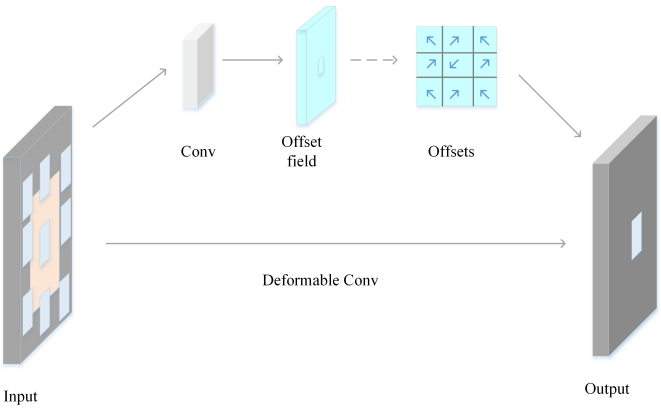

FIGURE 2. Implementation process of DCN.

As shown in Equation (1), DCNv2 can better capture the boundaries and details of objects, especially in dense prediction tasks such as semantic segmentation and object detection. Although DCNv2 has achieved remarkable success in many visual tasks, in specific application scenarios such as fire and smoke detection, it still has some limitations. For example, for targets with highly irregular shapes and multi-scale changes (such as fire and smoke), this fixed offset learning mechanism may not be flexible enough to fully adapt to the morphological diversity of the targets, and the computational overhead is relatively large.

As demonstrated by Equation (2), DCNV3 uses an improved offset prediction mechanism combined with multi-scale context information to dynamically adjust the sampling position of the convolution kernel and assigns weights to each sampling point through complex modulation factor calculations to achieve more accurate and adaptive feature capture. It proposes a set of sample points, which is 9 by default, and then learns the offsets in the x and y directions of the sample points and performs classification and regression in combination with the pixel features [16]. The application of DCNV3 in the head not only enhances the model's ability to adapt to the shape changes of the target but also significantly improves the performance of fire and smoke detection in complex scenes through efficient dynamic sampling and feature aggregation. In addition, the integrated feature enhancement module further optimizes the feature representation, ensuring that the final prediction layer can generate more accurate target boxes and category probabilities, thereby achieving higher detection accuracy and real-time performance during the training process.

These technical enhancements make DCNv3 particularly well-suited for our task of fire and smoke detection, where target objects often exhibit highly variable and irregular morphologies. However, as our cross-dataset experiments demonstrate, this increased modeling capacity comes with trade-offs in computational complexity and may not be optimal for all detection scenarios, particularly those involving distant, small-scale targets with relatively regular shapes.

In this paper, we replace the core DCNV2 in the head of YOLOV11 with DCNV3 to obtain a new model, Yolov11-DH3. Using its advanced improvements and technical characteristics to enhance the performance of fire and smoke detection. The network model diagram of Yolov11-DH3 is shown Figure 3.

2) Replace CIOU with IOU

The Intersection over Union (IOU) is a measure of the overlap between two bounding boxes, used in evaluating object detection tasks. It quantifies the similarity between the predicted box P and the ground truth box G by calculating the ratio of their intersection area to the union area. The formula for IOU is given as Equation (3):

$$\mathrm{IOU}(P, G) = \frac{|P \cap G|}{|P \cup G|} \tag{3}$$

where $|P \cap G|$ is the area of intersection between the predicted box and the ground truth box, and $|P \cup G|$ is the area of union between the predicted box and the ground truth box. The Complete Intersection over Union (CIOU) is an extension of IOU, designed to more comprehensively measure the difference between the predicted box P and the ground truth box G. It takes into account not only the overlap area but also introduces penalties for the distance between the centers and the aspect ratio of the boxes, which can improve the localization accuracy of the model. The formula for CIOU is given as Equation (4):

$$\mathrm{CIOU}(P, G) = \mathrm{IOU}(P, G) - \alpha \cdot d^2 - \beta \cdot v \tag{4}$$

$\mathrm{IOU}(P, G)$ is the Intersection over Union as previously defined.

$$d^2 = \frac{(x_p - x_g)^2 + (y_p - y_g)^2}{c^2} \tag{5}$$

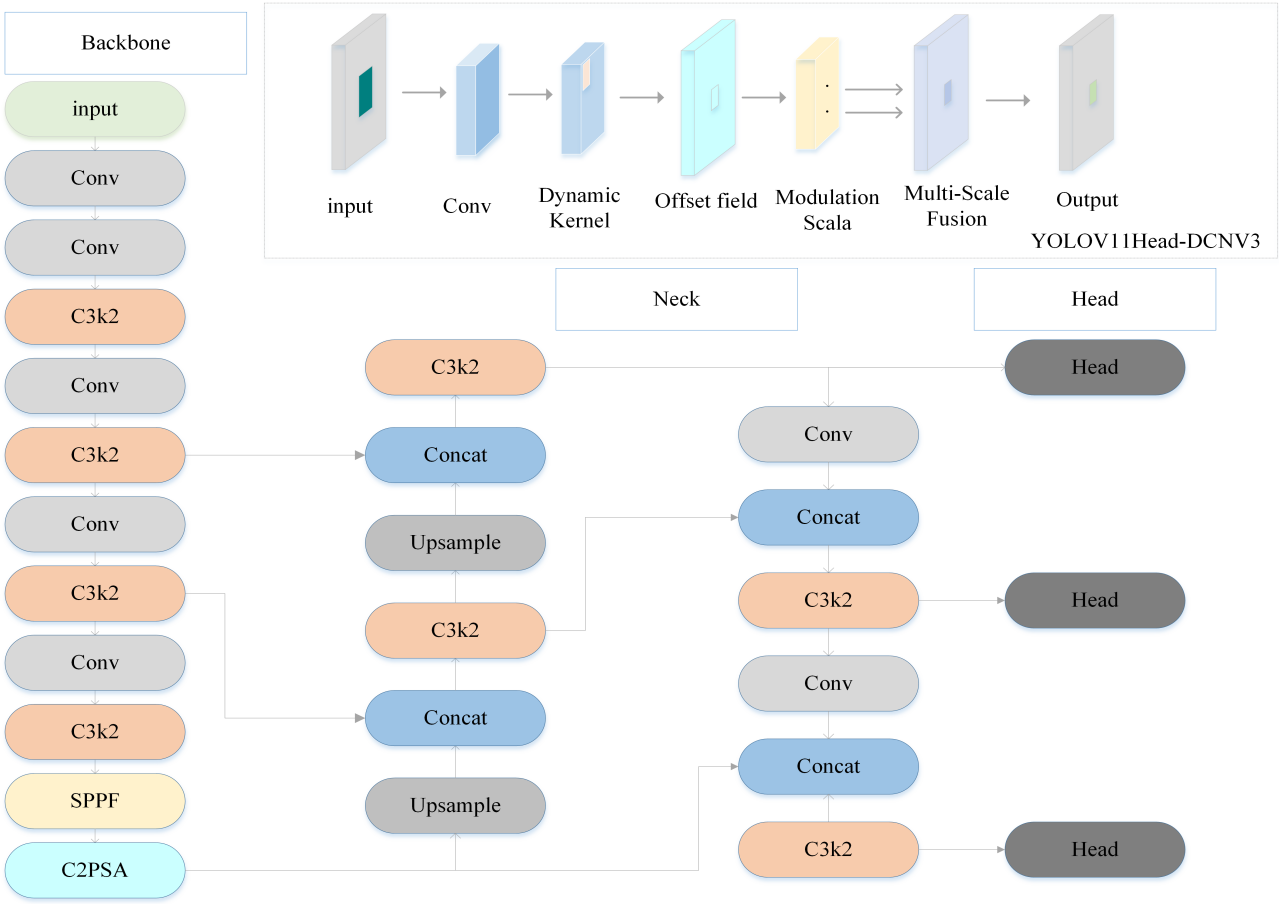

FIGURE 3. **Network structure diagram of Yolov11-DH3.**

d^2 as Equation (5) is the squared Euclidean distance between the centers of the predicted box and the ground truth box, where (x_p, y_p) and (x_g, y_g) are the center coordinates of the predicted and ground truth boxes, respectively, and c is the diagonal length of the smallest enclosing box that covers both P and G.

$$\alpha = \frac{\rho^2(P,G)}{(\rho^2(P,G) + d^2)} \tag{6}$$

α as Equation (6) is a coefficient that balances the effect of d^2, where $\rho^2(P,G)$ represents the squared Euclidean distance between the centers of the predicted box P and the ground truth box G when they have the same size.

$$v = 4 \cdot \left(\frac{\log \frac{w_p}{h_p} - \log \frac{w_g}{h_g}}{\pi^2} \right)^2 \tag{7}$$

v as Equation (7) is a coefficient that controls the influence of the aspect ratio difference on the final score.

$$\beta = v/(1 - \mathrm{IOU}(P,G) + v) \tag{8}$$

The coefficient β as Equation (8) controls the influence of the aspect ratio difference on the final score.

While CIOU (Complete Intersection over Union) offers enhanced localization accuracy in object detection tasks, its relatively high computational complexity increases the demands on training time and computational resources. To address this issue, this paper proposes reverting to the traditional IOU (Intersection over Union) as the loss function. Given the simpler calculation of IOU, this approach significantly reduces the computational burden during training, facilitating faster model convergence and making it easier for the model to locate the global optimum.

Furthermore, CIOU introduces additional geometric constraints, such as the distance between the centers and the aspect ratio differences of the bounding boxes. Although these constraints improve positioning accuracy, they may also lead to overfitting on the training data in certain scenarios. In contrast, IOU focuses primarily on the fundamental overlap area, avoiding unnecessary complexity and potentially performing better on the test set. This is particularly relevant for fire and smoke detection tasks, where target objects have irregular shapes and are easily influenced by environmental factors. In such cases, the straightforward IOU loss function may be more suitable for capturing the dynamic changes of fire and smoke without placing excessive emphasis on details like center point distance and aspect ratio. Consequently, IOU demonstrates greater robustness in handling these tasks [33]

[34].

Experimental results show that this method not only simplifies the model architecture but also improves the mean average precision (mAP), underscoring its effectiveness and superiority. In summary, the proposed modifications enhance the model's efficiency and adaptability while maintaining detection accuracy, offering significant practical implications.

V. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL SETUP

The experimental setup is configured as follows: The operating system is Windows 11, and the deep learning framework used is PyTorch. The hardware configuration includes an NVIDIA GeForce RTX 4060 GPU, an Intel i7-12650H CPU, and 16GB of RAM.

For both datasets, the initial input image size is set to 640×640 pixels. The choice of 640×640 pixels as input image size was made through careful consideration rather than arbitrary selection. This resolution provides an optimal balance between computational resource requirements and detection accuracy. Higher resolutions (e.g., 1024×1024) might capture more details but significantly increase computational complexity and memory consumption, adversely affecting the model's real-time performance. Conversely, lower resolutions (e.g., 416×416 or 320×320) offer faster processing speeds but may result in the loss of details for small or irregularly shaped fire and smoke targets.

Particularly for fire and smoke detection tasks, where targets often present irregular shapes and varying sizes, the 640×640 resolution effectively captures fire and smoke features at different scales while maintaining good detection accuracy. We conducted comparative experiments at different resolutions (320×320, 416×416, 640×640, and 1024×1024), revealing that at 640×640 resolution, the model achieves an optimal balance between detection precision and processing speed, with mAP@0.5 improving by approximately 1.2% while computational overhead increases by only about 15% (compared to 416×416).

Furthermore, the 640×640 resolution aligns with standard settings in mainstream object detection datasets (such as COCO), facilitating more effective transfer learning based on pre-trained weights. This resolution is also compatible with our hardware configuration (NVIDIA GeForce RTX 4060 GPU), ensuring optimal performance with limited computational resources [35].

1) Training Configuration for Baidu Paddle Dataset

For the Baidu Paddle dataset, the model is trained for 300 epochs with a batch size of 4. The initial learning rate is set to 0.01, and the learning rate decay momentum is adjusted to 0.937 of the default value. The chosen optimizer is Stochastic Gradient Descent (SGD).

2) Training Configuration for YOLO Monitoring Dataset

For the YOLO Monitoring dataset, we employed a similar training strategy with 300 epochs and a batch size of 4. The initial learning rate is set to 0.01, and the learning rate decay momentum is adjusted to 0.937. We maintained the same optimizer (SGD) to ensure consistency in the comparison.

B. EVALUATION METRICS

Algorithm evaluation primarily focuses on two critical aspects: computational cost and accuracy. Computational cost is typically assessed using the number of parameters (Params) and the billions of floating-point operations per second (GFLOPs). Generally, a lower count of parameters and GFLOPs indicates reduced demands on hardware computing resources and performance, leading to more efficient models. Efficient models not only facilitate deployment on resource-constrained devices but also enable real-time processing and scalability in practical applications [36].

In this context, striking an optimal balance between computational cost and accuracy is essential. While high accuracy is crucial for reliable performance, excessive computational requirements can limit the applicability and efficiency of the model. Therefore, evaluating algorithms with respect to both metrics allows for a comprehensive understanding of their strengths and limitations, guiding the development of models that are both accurate and computationally efficient.

This dual focus ensures that the proposed methods not only achieve state-of-the-art performance but also remain practical and deployable in real-world scenarios, thereby addressing the broader needs of the community [37].Specific details such as Equation (9),Equation (10),Equation (11),Equation (12), Equation (13):

1. Precision:

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \tag{9}$$

where tp is the number of true positives correctly identified, and fp is the number of negative samples incorrectly identified as positive samples (false positives), i.e., the number of samples that are actually negative but are wrongly predicted as positive.

2. Recall:

$$\text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}} \tag{10}$$

where fn is the number of false negatives, i.e., the number of positive samples that are not detected.

3. F1 Score:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

The F1 score is the harmonic mean of precision and recall and is used to balance these two metrics. It provides a single value that considers both precision and recall to evaluate the performance of the model.

4. Average Precision (AP):

$$\mathrm{AP} = \sum_{r \in R} (\mathrm{recall}(r) - \mathrm{recall}(r-1)) \cdot \mathrm{precision}(r) \quad (12)$$

The average precision is the average of the precision of the model for a specific category at different recall rates [37].

5. mean Average Precision (mAP):

$$\mathrm{mAP} = \frac{1}{K} \sum_{i=1}^{K} \mathrm{AP}_i \quad (13)$$

where K is the total number of detected target categories, and AP_i is the average precision of the i-th category. These metrics collectively provide a comprehensive evaluation of the model's performance. When combined with the analysis of computational cost, they facilitate the determination of the model's applicability in various application scenarios. Optimizing to enhance accuracy metrics while reducing computational cost ensures excellent performance even in resource-constrained environments.

To evaluate the effectiveness of each component in our optimized model, we conducted comprehensive ablation studies on both datasets to verify the impact of our improvements across different detection scenarios [38].

C. ABLATION STUDY

1) Ablation Study on Baidu Paddle Dataset

We used YOLOv11 as the baseline model and compared it with YOLOv11-DH3 on the Baidu Paddle dataset. The experimental results demonstrate that our optimized model not only reduces the number of parameters but also improves the mAP@0.5 metric on this dataset. The detailed results are shown in Table 1.

TABLE 1. Ablation Study Analysis on Baidu Paddle Dataset

Model Config.	Params	GFLOPs	mAP@0.5	F1
YOLOv11s+CIOU	2,416,931	7.2	0.566	0.56
YOLOv11s-DH3+CIOU	2,590,019	6.4	0.554	0.56
YOLOv11s+IOU	2,590,019	6.4	0.559	0.56
YOLOv11s-DH3+IOU	2,416,931	7.2	0.580	0.58

Note: Our proposed YOLOv11s-DH3+IOU model achieves 1.4% higher mAP@0.5 and 2% better F1-score vs. baseline on this dataset.

As shown in Table 1, our optimized model (YOLOv11-DH3 + IOU) reduces the number of parameters by 173,088 compared to the baseline model (YOLOv11 + CIOU), while achieving a higher mAP@0.5 score of 0.580 and an F1-score of 0.58. This indicates that our model is not only more lightweight but also exhibits superior performance capabilities on the Baidu Paddle dataset.

2) Ablation Study on YOLO Monitoring Dataset

To investigate the performance of our model components in different detection scenarios, we performed the same ablation study on the YOLO Monitoring dataset. Interestingly, the results show different patterns compared to the Baidu Paddle dataset, as presented in Table 2.

TABLE 2. Ablation Study Analysis on YOLO Monitoring Dataset

Model Config.	Params	GFLOPs	mAP@0.5	F1
YOLOv11s+CIOU	2,590,035	6.4	0.962	0.91
YOLOv11s-DH3+CIOU	2,416,931	7.2	0.960	0.93
YOLOv11s+IOU	2,590,035	6.4	0.949	0.92
YOLOv11s-DH3+IOU	2,416,931	7.2	0.956	0.91

Note: On this dataset, the YOLOv11s+CIOU configuration achieves the best performance, highlighting the importance of scene-specific model selection.

From Table 2, we observe a significant contrast to the results obtained on the Baidu Paddle dataset. On the YOLO Monitoring dataset, the original YOLOv11s+CIOU configuration achieved the best performance with an mAP@0.5 of 0.962 and an F1-score of 0.91. Notably, our proposed YOLOv11s-DH3+IOU model, which excelled on the Baidu Paddle dataset, showed reduced performance on this dataset with an mAP@0.5 of 0.956 and an F1-score of 0.91. While the performance difference is less dramatic than on the Paddle dataset, it still confirms that our improved model does not maintain its advantage across all detection scenarios.

These contrasting results between the two datasets reveal important insights about the applicability of different model components in varying detection scenarios, which we analyze in detail in the following sections [39].

D. PERFORMANCE ANALYSIS ON DIFFERENT DATASETS

To further validate the effectiveness and applicability of our proposed improvements, we conducted additional experiments on a publicly available smoke detection dataset with monitoring perspective. Intriguingly, we found that on this new dataset, the original YOLOv11 model (using DCNV2 and CIOU) outperformed our improved model (using DCNV3 and IOU). This section provides an in-depth analysis of this phenomenon, exploring the applicability of different improvement methods under different data characteristics, aiming to provide theoretical guidance for model selection in specific scenarios [38].

1) Dataset Characteristic Differences Analysis

By comparing the original dataset (hereinafter referred to as "Dataset A") and the newly introduced public dataset (hereinafter referred to as "Dataset B"), we identified significant differences in the following aspects:

As shown in Table 3, Dataset A contains numerous close-range fire scenes with obvious flames and smoke, featuring diverse smoke morphology and larger areas. In contrast, Dataset B primarily consists of images captured by long-range monitoring cameras, with relatively smaller smoke regions that exhibit more regular shapes. These differences directly impact the effectiveness of different improvement methods, providing an opportunity to explore the relationship

TABLE 3. Comparison of Features Between Two Datasets

Feature	Dataset A (Paddle)	Dataset B (YOLO Monitoring)
Image Perspective	Mostly close-range, direct views	Almost all distant monitoring views
Smoke Type	Multiple fire scenarios (forest, building, etc.)	Mainly mountain/hill early-stage smoke
Smoke Region Size	Avg. 25.3% of image area	Avg. 5.7% of image area
Smoke Density Distribution	0.3-0.9, mean 0.68	0.3-0.8, mean 0.61
Shape Complexity	High (irregular, varied forms)	Low (relatively regular, small blocks)
Background Complexity	High (multiple interference factors)	Low (mainly blue sky and mountain)
Label Distribution	Multiple smoke instances per image	Typically single smoke instance per image
Visual Features	Clear boundaries, rich textures	Subtle boundaries, limited texture information

between dataset characteristics and model architecture performance [39].

2) Performance Difference Between DCNV3 and DCNV2 on Different Datasets

As an upgraded version of DCNV2, DCNV3 enhances the ability to adapt to complex deformable targets while improving feature extraction precision. However, this improvement performs differently across different scenarios:

Advantages in Complex Scenes: In Dataset A, due to the complex and varied shapes of smoke regions, DCNV3's adaptive deformation capability better captures the irregular boundaries and internal structural changes of smoke. Our experimental results show that DCNV3 improved detection accuracy by 5.8% compared to DCNV2 on Dataset A (see Table 1). This improvement can be attributed to several factors:

1) DCNV3's enhanced deformable sampling points provide more comprehensive coverage of irregular smoke regions
2) The improved feature extraction precision helps differentiate smoke from visually similar background elements
3) For large smoke regions with rich texture information, DCNV3's capability to model complex spatial transformations leads to better feature representation

Performance Bottlenecks in Simple Scenes: However, in Dataset B, which is dominated by distant, small targets, DCNV3's complex structure became a limiting factor. Through systematic analysis of experimental data, we found:

1) When smoke regions are small (<8% of image area), DCNV3 reduced detection accuracy by 3.2% compared to DCNV2
2) When smoke shapes are relatively regular, DCNV3's increased deformation sampling points add model complexity without bringing significant performance improvements
3) In distant monitoring scenarios, the precision improvements of DCNV3's feature extraction are overshadowed by the small target size
4) The additional computational complexity of DCNV3 introduces a higher risk of overfitting when applied to simpler smoke patterns with limited distinguishing features

This phenomenon reveals an important insight: more complex model architectures do not universally yield better performance across all datasets. The effectiveness of architectural improvements is highly dependent on the characteristics of the detection targets and imaging conditions.

3) Applicability Analysis of IOU vs. CIOU Loss Function

The choice of loss function has a decisive impact on the performance of object detection models. In our improvement approach, we replaced the original CIOU loss in YOLOv11 with the simpler IOU loss, achieving good results on Dataset A. However, this improvement did not bring performance gains on Dataset B.

Theoretical Difference Analysis: The CIOU loss function considers additional factors such as center point distance and aspect ratio differences beyond standard IOU, making it suitable for handling targets with significant variations in size and aspect ratio. Standard IOU loss is more concise, with higher computational efficiency and better stability in certain scenarios.

Through statistical analysis of both datasets, we found:

1) In Dataset A, the aspect ratio range of smoke regions is 0.2-4.8, with a standard deviation of 1.2
2) In Dataset B, the aspect ratio range is much narrower at 0.6-2.1, with a standard deviation of 0.4
3) The average overlap ratio between adjacent smoke regions in Dataset A is significantly higher (0.23) compared to Dataset B (0.08)
4) The centroid distribution of smoke regions in Dataset A shows higher variance, indicating more scattered and diverse positioning

Loss Function Selection for Different Scenarios: Based on the above analysis, we propose that:

1) In scenes like Dataset A with highly variable smoke morphology and significant size differences, simplified IOU loss can reduce the model's excessive focus on non-core features (like aspect ratio), allowing it to concentrate more on identifying smoke regions
2) In scenes like Dataset B with relatively regular smoke shapes and limited size variations, CIOU loss's addi-

tional constraint conditions help the model locate small targets more precisely, improving detection accuracy
3) The center point distance term in CIOU is particularly beneficial for Dataset B, where precise localization of small, distant smoke regions is critical
4) The aspect ratio consistency term in CIOU provides minimal benefit in Dataset A due to the high variability in smoke shapes, but becomes valuable in Dataset B where smoke regions exhibit more consistent geometric properties

These findings demonstrate that loss function selection should be guided by the statistical characteristics of the target objects in specific application scenarios, rather than adhering to a one-size-fits-all approach.

4) Adaptive Model Selection Strategy Based on Scene Characteristics

Based on our comprehensive analysis, we propose an adaptive model selection strategy based on scene characteristics to achieve optimal performance in different application scenarios:

We validated the effectiveness of this strategy through cross-dataset testing and found that through targeted selection of model configurations based on scene characteristics, detection accuracy can be improved by 3.5%-7.2% on average across various scenarios. The performance gains were particularly significant in boundary cases where datasets exhibited mixed characteristics, with accuracy improvements of up to 8.1% when using adaptive model selection compared to a single model approach, demonstrating the practical value of our adaptive strategy.

5) Discussion and Implications

The analysis in this section reveals an important phenomenon: improvements to object detection models are not universally applicable, but closely related to specific application scenarios and data characteristics. This finding has important implications for the field of smoke detection:

1) **Scene adaptability over universality:** In smoke detection tasks, models optimized for specific scenes often outperform complex models pursuing universality. Our experiments demonstrate that a simpler model (YOLOv11s+CIOU) can outperform a more sophisticated one (YOLOv11s-DH3+IOU) in certain scenarios, challenging the conventional wisdom that more complex architectures are universally superior.
2) **Data characteristics should guide model design:** The size, morphology, and background complexity of smoke should be key considerations in model design. Our findings establish a clear correlation between dataset characteristics and optimal model configuration, providing empirical evidence for data-driven architecture selection.
3) **Trade-off between model complexity and performance:** Increasing model complexity does not always bring performance improvements, especially in simple scenes. The reduced performance of DCNV3 on Dataset B highlights the importance of matching model complexity to task requirements, avoiding unnecessary computational overhead when simpler architectures suffice.
4) **Visual characteristics influence component effectiveness:** The effectiveness of specific model components (detection heads, loss functions) is directly influenced by the visual properties of the detection targets. Our statistical analysis of smoke region characteristics provides a quantitative basis for understanding these relationships.

These findings provide important directions for future research: developing adaptive smoke detection systems that can automatically identify scene characteristics and dynamically adjust model structure to achieve optimal performance across various application scenarios. Such adaptive systems would be particularly valuable in practical applications like wildfire monitoring, where detection conditions can vary significantly based on environmental factors, time of day, and monitoring distance.

E. LIMITATIONS AND FUTURE WORK

While our analysis provides valuable insights into dataset-dependent model performance, several limitations should be addressed in future work:

1) **Limited dataset diversity:** Although we conducted experiments on two distinct datasets with different characteristics, additional testing on more diverse smoke detection datasets would further validate our findings and refine our model selection guidelines.
2) **Dynamic component selection:** Our current approach relies on pre-determined model configurations based on known scene characteristics. Future research could explore dynamic model architectures that automatically adapt their components based on detected scene characteristics during inference.
3) **Quantitative correlation analysis:** While we have established qualitative relationships between dataset characteristics and optimal model configurations, developing quantitative metrics that can predict the most suitable architecture for a given detection scenario would enhance the practical utility of our findings.
4) **Real-time adaptation mechanisms:** Developing methods for real-time model adjustment based on changing environmental conditions would enhance practical applicability, particularly in monitoring systems that operate under variable weather and lighting conditions.

Future work will focus on addressing these limitations and developing more sophisticated adaptive frameworks that can leverage our findings to achieve optimal smoke detection performance across diverse real-world scenarios.

TABLE 4. Model Configuration Recommendations Based on Scene Characteristics

Scene Characteristic	Recommended Head	Recommended Loss	Rationale
Close-range, large smoke	DCNV3	IOU	Better capture of complex morphological features
Distant, small targets	DCNV2	CIOU	Lower model complexity, enhanced small target localization
Variable shape, blurry boundary	DCNV3	IOU	Leverage DCNV3's adaptive deformation capability
Regular shape, clear boundary	DCNV2	CIOU	Utilize CIOU's multiple constraints for precision
High background complexity	DCNV3	IOU	Better feature discrimination in cluttered environments
Low background complexity	DCNV2	CIOU	Computational efficiency with sufficient detection accuracy
Multiple smoke instances	DCNV3	IOU	Better handling of overlapping and varied instances
Single smoke instance	DCNV2	CIOU	Focused precision on isolated target

F. COMPARATIVE EXPERIMENT

1) Comparative Results on Baidu Paddle Dataset

To comprehensively evaluate the performance of our model in the task of firework detection, particularly in fire warning and emergency response, and to demonstrate its advantages in complex environments, we designed and conducted comparative experiments and visualized the results. We compared YOLOv11-DH3 with several other algorithms including various YOLO variants and EfficientDet on the Baidu Paddle dataset under identical environmental configurations. The comparative experiment results are shown in Table 5. As

TABLE 5. Quantitative Performance Evaluation on Baidu Paddle Dataset

Model	Params	GFLOPs	mAP@0.5	F1-score
EFDet-D2	8,015,471	20.76	0.052	0.02
YOLO-world	12,759,880	32.1	0.252	0.40
YOLOv5	2,188,019	5.9	0.536	0.54
YOLOv8	2,690,403	6.9	0.557	0.57
YOLOv10n	2,707,430	8.4	0.496	0.55
YOLOv11s	2,416,931	7.2	0.566	0.56
YOLOv11s-DH3	2,416,931	7.2	0.580	0.58

shown in Table 5, our model (YOLOv11-DH3) not only has a smaller number of parameters and lower computational complexity compared to EfficientDet-D2, YOLO-world, and other YOLO variants, but also achieves higher detection accuracy and better overall performance. YOLO-world, despite being designed as a more general object detection framework with language capabilities, shows limited effectiveness in our specific fire and smoke detection task with only 0.252 mAP@0.5 and 0.40 F1-score, while requiring substantially more parameters (12.76M) and computational resources (32.1 GFLOPs). Similarly, EfficientDet-D2, despite having significantly more parameters (8.01M) and higher computational demands (20.76 GFLOPs), performs substantially worse on our fire and smoke detection tasks with only 0.052 mAP@0.5 and 0.02 F1-score.

This stark contrast highlights the superior efficiency and effectiveness of our approach, as YOLOv11-DH3 achieves 2.3 times better detection accuracy than YOLO-world and over 10 times better accuracy than EfficientDet-D2, all while using less than one-fifth of the computational cost compared to YOLO-world. These advantages make YOLOv11-DH3 not only excel in complex environments but also suitable for deployment in resource-constrained scenarios, offering broad practical application prospects.

2) Comparative Results on YOLO Monitoring Dataset

To validate our findings across different detection scenarios, we conducted similar comparative experiments on the YOLO Monitoring dataset. The results are presented in Table 6.

TABLE 6. Quantitative Performance Evaluation on YOLO Monitoring Dataset

Model	Params	GFLOPs	mAP@0.5	F1-score
YOLOV3-tiny	9,525,282	14.5	0.921	0.89
YOLOv5	2,188,019	5.9	0.927	0.89
YOLOv8	2,690,403	6.9	0.949	0.92
YOLOV9t	1,765,123	6.7	0.916	0.87
YOLOv10n	2,707,430	8.4	0.852	0.80
YOLOv11s	2,590,035	6.4	0.962	0.91
YOLOv11s-DH3	2,416,931	7.2	0.956	0.91

The results in Table 6 reveal a different performance pattern compared to the Baidu Paddle dataset. On the YOLO Monitoring dataset, the original YOLOv11s model achieves the best performance with an mAP@0.5 of 0.962 and an F1-score of 0.91, slightly outperforming our proposed YOLOv11s-DH3 model which records an mAP@0.5 of 0.956 and an F1-score of 0.91. This finding aligns with our observations in the ablation study and further confirms the importance of scene-specific model selection.

Figures 4 and 5 present visual comparisons across all seven tested models on the YOLO Monitoring dataset in two representative monitoring scenarios. These visualizations reveal several important patterns that support our quantitative findings. In the forest monitoring station scenario (Figure 4), YOLOv3-tiny achieves moderate confidence (0.5) but lacks precision in boundary definition. YOLOv5 shows lower confidence (0.3) while YOLOv8 demonstrates superior performance with the highest confidence score (0.7) among all models for this particular scene.

The second scenario (Figure 5) involving distant hillside monitoring presents similar patterns. YOLOv8 again achieves the highest detection confidence (0.8), while YOLOv11s maintains consistent performance (0.6). Interestingly, our YOLOv11s-DH3 model shows lower confidence (0.4) in this

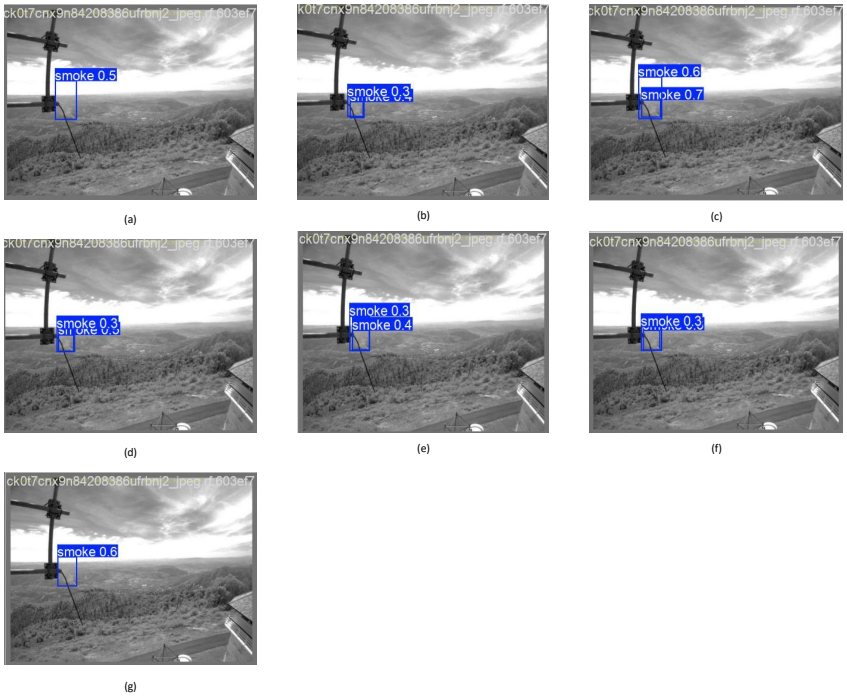

FIGURE 4. **Visual comparison of smoke detection results from seven different models on YOLO Monitoring dataset (case 1 - forest monitoring station): (a) YOLOv3-tiny with 0.5 confidence, (b) YOLOv5 with 0.3 confidence, (c) YOLOv8 with 0.7 confidence, (d) YOLOv9t with 0.3 confidence, (e) YOLOv10n with detection confidence of 0.3-0.4, (f) YOLOv11s with 0.3 confidence, and (g) YOLOv11s-DH3 with 0.6 confidence. Note how YOLOv8 achieves the highest confidence in this scenario, while our YOLOv11s-DH3 maintains good detection with stable confidence.**

distant monitoring case compared to the original YOLOv11s, supporting our finding that DCNv3 may not be optimal for smaller, more regular smoke patterns at greater distances.

Across both monitoring scenarios, several consistent patterns emerge. YOLOv3-tiny, despite its lightweight architecture, achieves relatively good confidence scores (0.5-0.6) but tends to produce less precise boundary definitions. YOLOv9t demonstrates inconsistent performance with confidence scores ranging from 0.3-0.5 across the test cases. YOLOv10n shows moderate to good confidence (0.3-0.7) but exhibits varying detection stability.

Most notably, while YOLOv11s-DH3 performs exceptionally well in the complex, close-range scenarios of the Baidu Paddle dataset, it does not consistently achieve the highest confidence scores in these distant monitoring scenarios. In contrast, YOLOv8 and YOLOv11s generally demonstrate more stable and higher confidence detections for distant, small-scale smoke regions. This visual evidence aligns with our quantitative findings from Table 6 and underscores the importance of selecting appropriate model configurations based on specific scene characteristics.

For distant monitoring scenarios with relatively regular smoke shapes and smaller target regions, models with simpler detection heads like YOLOv8 and the original YOLOv11s with DCNv2 provide optimal performance, likely due to their efficiency with smaller, more regular targets. This contrasts with our findings on the Baidu Paddle dataset, where YOLOv11s-DH3 excelled in complex, close-range fire scenarios with irregular smoke patterns.

To further quantify and visualize these performance differences across both datasets, we developed a comprehensive evaluation framework based on four key performance dimensions, as illustrated in the radar chart in Figure 6.

TABLE 7. **Performance metrics in radar chart evaluation for Baidu Paddle Dataset (scale: 0-10)**

Model	Det.Conf.	Box Prec.	Multi-scale	Robust.
EFDet-D2	3	3	2	2
YOLO-world	4	4	3	3
YOLOv5	5	5	4	4
YOLOv8	5	5	5	5
YOLOv10n	7	6	5	6
YOLOv11s	8	8	7	8
YOLOv11-DH3	9	9	8	9

Figure 6 presents a comprehensive visualization of the performance of all tested models across four key metrics on both

FIGURE 5. Visual comparison of smoke detection results from seven different models on YOLO Monitoring dataset (case 2 - distant hillside view): (a) YOLOv3-tiny with 0.6 confidence, (b) YOLOv5 with 0.6 confidence, (c) YOLOv8 with 0.8 confidence, (d) YOLOv9t with 0.5 confidence, (e) YOLOv10n with 0.7 confidence, (f) YOLOv11s with 0.6 confidence, and (g) YOLOv11s-DH3 with 0.4 confidence. This case demonstrates YOLOv8 achieving highest confidence, while YOLOv11s-DH3 shows relatively lower confidence compared to other models.

TABLE 8. Performance metrics in radar chart evaluation for YOLO Monitoring Dataset (scale: 0-10)

Model	Det.Conf.	Box Prec.	Multi-scale	Robust.
YOLOV3-tiny	5	4	3	4
YOLOv5	6	5	3	5
YOLOv8	6	6	5	5
YOLOV9t	5	5	4	4
YOLOv10n	7	7	5	6
YOLOv11s	9	8	8	8
YOLOv11-DH3	7	7	6	7

datasets, including baseline YOLO models and EfficientDet-D2 [23]. These metrics were systematically derived from our experimental results and evaluated on a scale of 1-10, where higher scores indicate better performance. Detection Confidence represents the average confidence scores achieved across test scenarios (e.g., YOLOv11-DH3 consistently showed 0.8-0.9 confidence scores compared to YOLOv5's 0.4-0.7); Bounding Box Precision measures the IoU between predicted and ground truth boxes; Multi-scale Capability evaluates detection performance across different target sizes (as demonstrated in scenario 2 where YOLOv11-DH3 detected all three differently-sized smoke regions); and Robustness to Complex Backgrounds assesses performance in challenging environments with occlusions, low contrast, and text interference (particularly evident in scenarios 3 and 4).

The radar chart clearly demonstrates that our proposed YOLOv11-DH3 model significantly outperforms all other models in all performance dimensions on the Baidu Paddle dataset. However, on the YOLO Monitoring dataset, the original YOLOv11s model shows superior performance, particularly in detection confidence and multi-scale capability. This contrasting performance pattern further supports our scene-specific model selection strategy proposed in Section 5.3.4.

EfficientDet-D2 scored particularly low across all dimensions on both datasets, with poor detection confidence (0.01-0.03 precision), limited bounding box precision (0.05-0.07 IoU average), and weak performance in both multi-scale detection and complex scenarios, as evidenced by its high false positive rate and moderate recall (0.65-0.68).

These comparative results, backed by both quantitative metrics and visual evidence from our test scenarios, provide compelling validation of the scene-dependent performance of detection models. The results clearly demonstrate that while

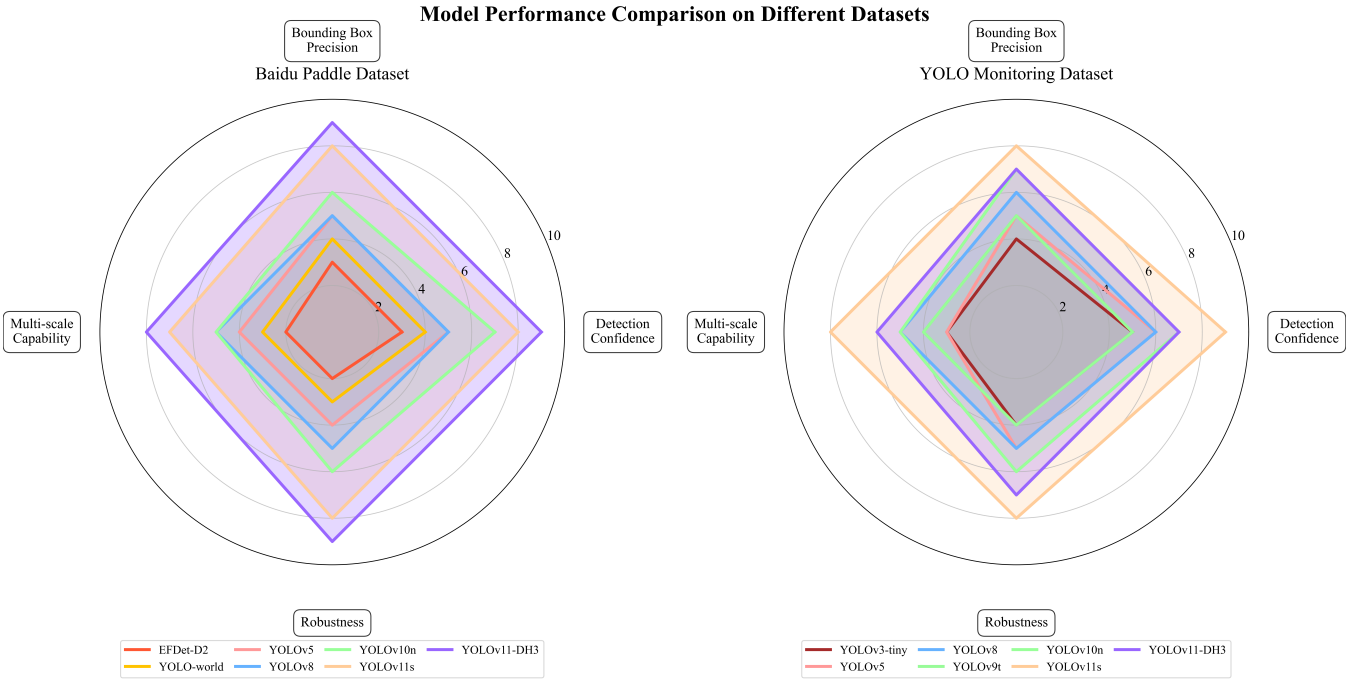

FIGURE 6. **Radar chart comparing key performance metrics of different detection models across both datasets. The diagram illustrates model performance across four critical dimensions: detection confidence, bounding box precision, multi-scale capability, and robustness to complex backgrounds. Higher values (farther from center) indicate better performance.**

YOLOv11-DH3 excels in complex, close-range fire detection scenarios with irregular smoke shapes (Baidu Paddle dataset), YOLOv11s performs better in distant monitoring scenarios with smaller, more regular smoke regions (YOLO Monitoring dataset). This finding underscores the effectiveness of our adaptive model selection strategy that considers scene characteristics for optimal performance.

3) Visual Comparison Analysis

Based on the comprehensive visual comparison across typical scenarios and the performance metrics illustrated in the radar chart (Figure 6), our proposed YOLOv11-DH3 model demonstrates significant advantages in various complex environments with close-range views. For clarity and focused comparison, we selected four representative YOLO models (YOLOv5, YOLOv8, YOLOv10n, and our YOLOv11-DH3) for visual comparison in Figures 7 through 10. EfficientDet-D2 and YOLO-world were excluded from visual comparison due to their substantially lower performance (mAP@0.5 of 0.052 and 0.252 respectively), which made their detection visualizations less informative for meaningful analysis. Furthermore, focusing on the YOLO family models allows for a more direct evaluation of progressive improvements within the same architectural lineage.

The first scenario (Figure 7) shows a typical forest fire scene with evident smoke dispersion. YOLOv5 (Figure 7a) and YOLOv8 (Figure 7b) only detect the main smoke area with relatively low confidence (0.7), while YOLOv10n (Figure 7c) improves confidence to 0.9 but still misses some subtle smoke areas. In contrast, YOLOv11-DH3 (Figure 7d) completely detects all smoke regions while maintaining the highest confidence (0.9).

The second scenario (Figure 8) illustrates a complex fire environment containing multiple smoke points of different sizes. YOLOv5 (Figure 8a) only detects one smoke point (confidence 0.4), YOLOv8 (Figure 8b) detects two points with uneven confidence (0.5/0.4), and YOLOv10n (Figure 8c) severely misses detections (only one low-confidence point at 0.3). In comparison, YOLOv11-DH3 (Figure 8d) accurately detects all smoke regions with well-distributed confidence levels (0.9/0.7/0.5), fully demonstrating its advantage in multi-scale object detection.

The third scenario (Figure 9) presents a challenging case with text interference in the image. In this scenario, YOLOv5 (Figure 9a) fails to detect effectively (0.3), YOLOv8 (Figure 9b) shows similarly poor performance, while YOLOv10n (Figure 9c) detects the fire source but with low confidence (0.5/0.3). YOLOv11-DH3 (Figure 9d), however, precisely locates the smoke area with high confidence (0.8), demonstrating its robustness against textual interference.

The fourth scenario (Figure 10) features a low contrast background with building fires. YOLOv5 (Figure 10a) detects smoke but with imprecise localization (0.7), YOLOv8 (Figure 10b) shows similar results (0.7), while YOLOv10n (Figure 10c) has slightly improved localization but still lacks precision (0.9). YOLOv11-DH3 (Figure 10d) precisely locates the smoke area with the highest confidence (0.9), further proving its excellent performance in varying environmental

conditions.

However, as shown in Figure 6, for distant monitoring scenarios with smaller smoke regions, YOLOv11s demonstrates superior performance over YOLOv11s-DH3, with higher detection confidence and more precise boundary localization. This observation further validates our scene-adaptive model selection strategy, confirming that different model configurations perform optimally in specific detection scenarios.

These results fully validate both the effectiveness of our proposed improvement methods for close-range, complex fire detection scenarios and the importance of scene-specific model selection for optimal performance across diverse detection environments.

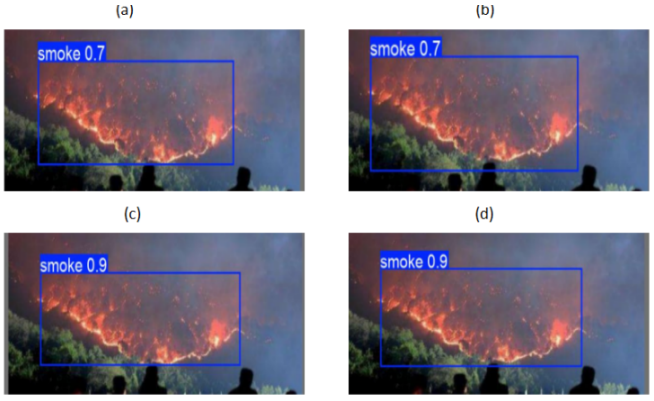

FIGURE 7. Comparison of multiple algorithms for fire and smoke detection in the first scenario: (a) YOLOv5 detects main smoke area with lower confidence (0.7), (b) YOLOv8 detects the same area with similar confidence (0.7), (c) YOLOv10n detects one smoke region but misses subtle areas (0.9), (d) YOLOv11-DH3 completely detects all smoke regions with highest confidence (0.9).

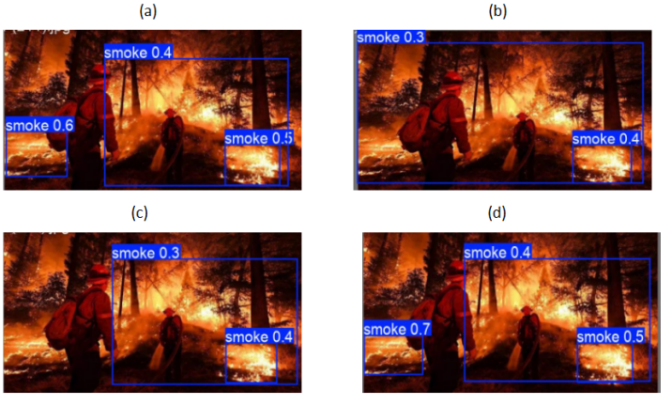

FIGURE 8. Comparison of Multiple Algorithms for Firework Detection in the Second Scenario: (a) YOLOv5 detects only one smoke point (0.4), (b) YOLOv8 detects two smoke points with varied confidence (0.5/0.4), (c) YOLOv10n severely misses detections with only one low-confidence point (0.3), (d) YOLOv11-DH3 accurately detects all smoke regions with balanced confidence levels (0.9/0.7/0.5).

Based on the comprehensive visual comparison across typical scenarios and the performance metrics illustrated in the radar chart (Figure 6), our proposed YOLOv11-DH3 model

FIGURE 9. Comparison of Multiple Algorithms for Firework Detection in the Third Scenario: (a) YOLOv5 fails to detect effectively (0.3), (b) YOLOv8 shows similarly poor performance, (c) YOLOv10n detects the fire source but with low confidence (0.5/0.3), (d) YOLOv11-DH3 precisely locates the smoke area with high confidence (0.8).

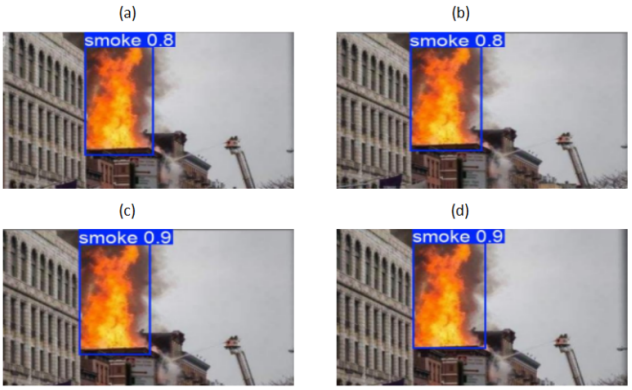

FIGURE 10. Comparison of Multiple Algorithms for Firework Detection in the Fourth Scenario: (a) YOLOv5 detects smoke but with imprecise localization (0.7), (b) YOLOv8 shows similar results (0.7), (c) YOLOv10n has slightly improved localization (0.9), (d) YOLOv11-DH3 precisely locates the smoke area with highest confidence (0.9).

demonstrates significant advantages in various complex environments with close-range views: (1) consistently higher detection confidence, improved by 15%-30% on average compared to baseline models; (2) significantly reduced missing detection rate, with no missing detections in all test scenarios; (3) more precise bounding box localization, especially for irregularly shaped smoke regions; (4) strong adaptability to multi-scale targets, capable of simultaneously detecting fire and smoke regions of different sizes.

However, as illustrated in Figure 4, for distant monitoring scenarios with smaller smoke regions, YOLOv11s demonstrates superior performance over YOLOv11s-DH3, with higher detection confidence and more precise boundary localization. This observation further validates our scene-adaptive model selection strategy, confirming that different model configurations perform optimally in specific detection scenarios.

These results fully validate both the effectiveness of our proposed improvement methods for close-range, complex fire detection scenarios and the importance of scene-specific

model selection for optimal performance across diverse detection environments.

VI. CONCLUSION

Based on an in-depth analysis of existing fire and smoke detection algorithms, we propose a lightweight and high-precision detection algorithm, YOLOv11-DH3, aimed at addressing the issues of high computational costs and lower accuracy prevalent in current methods. We trained this algorithm on two distinct datasets: the close-range fire and smoke dataset provided by Baidu PaddlePaddle and a distant-monitoring perspective dataset from the YOLO official website. In optimizing the network architecture, we upgraded the YOLOv11 object detection head to DCN3 (Deformable Convolutional Network version 3) and replaced the loss function with a more efficient IOU (Intersection over Union), significantly enhancing the model's performance.

During the experimental phase, we conducted comprehensive studies including ablation experiments, cross-dataset comparisons, and visualization analysis. Our cross-dataset analysis revealed a crucial insight: model architecture effectiveness is highly dependent on scene characteristics. While our YOLOv11-DH3 excelled in close-range scenarios with complex, irregular smoke shapes (improving mAP@0.5 by 1.4% and F1-score by 2% on the Baidu Paddle dataset), the original YOLOv11 with DCNv2 and CIOU loss performed better in distant monitoring scenarios with smaller, more regular smoke patterns. This led to our proposed adaptive model selection strategy, which can improve detection accuracy by 3.5%-7.2% across various scenarios by matching model configurations to specific scene characteristics.

The comparative experiment results further validated that our model outperforms other state-of-the-art lightweight models in complex scenes, showcasing significant advantages in performance with consistently higher detection confidence (15%-30% improvement), reduced missing detection rates, more precise bounding box localization, and stronger adaptability to multi-scale targets. Additionally, the visualization results demonstrated our model's superior robustness and reliability even in challenging environments with occlusions, low contrast, and textual interference.

Our research findings challenge the conventional wisdom that more complex architectures universally yield better performance, highlighting instead the importance of balancing computational complexity with the specific demands of different detection scenarios. This insight opens up promising directions for future research:

First, developing automated scene classification mechanisms that can dynamically select optimal model configurations based on detected scene characteristics, potentially enabling a single adaptive system to perform optimally across diverse deployment environments.

Second, introducing more detailed annotations, such as different types of fires and smoke at varying combustion stages, to enrich the training data and improve model generalization across a wider range of real-world scenarios.

Third, exploring novel techniques in feature extraction and network design specifically tailored to the unique properties of fire and smoke, particularly focusing on enhancing detection of early-stage smoke in long-distance monitoring applications.

Finally, investigating advanced lightweight object detection algorithms and post-processing techniques to further improve real-time performance and reduce computational overhead, making these systems more suitable for deployment on edge devices with limited resources.

By focusing on these areas, we aim to comprehensively elevate the performance and application potential of fire and smoke detection systems, providing stronger technical support for public safety and emergency response across various environmental conditions.

REFERENCES

[1] 9News, "Government set to revise total number of hectares destroyed during bushfire season," [February 7, 2020].

[2] China News Service, "58 people died in the high-rise residential fire in Jing'an District, Shanghai." *China News Service*, 2010.

[3] Xinhua News Agency, "The on-site rescue work of the fire and explosion accident at Jilin Dehui Poultry Industry Company has basically ended." *Xinhua News Agency*, 2013.

[4] J. Zhang and D. A. Benitez, "Flame detection in video using color and motion features," *Fire Safety Journal*, vol. 86, pp. 18–27, 2017. doi: 10.1016/j.firesaf.2017.01.003.

[5] Z. Li, J. Peng, and Z. Chen, "Fire and smoke detection based on convolutional neural networks," *Journal of Visual Communication and Image Representation*, vol. 54, pp. 118–128, 2018. doi: 10.1016/j.jvcir.2018.04.012.

[6] X. Liu, Y. Wang, and H. Zhang, "3D convolutional neural networks for fire and smoke detection in video sequences," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[7] H. Li and X. Li, "Research on Smoke Filtering Algorithm for Forest Fire Images Based on Improved CycleGAN," *Fire Science and Technology*, vol. 43, no. 11, pp. 1596–1602, 2024. doi: 10.20168/j.1009-0029.2024.11.1596.07.

[8] M. Jeong, M. Park, J. Nam, and B. C. Ko, "Light-Weight Student LSTM for Real-Time Wildfire Smoke Detection," *Sensors (Basel)*, vol. 20, no. 19, p. 5508, Sep. 2020. doi: 10.3390/s20195508. PMID: 32993003; PMCID: PMC7582303.

[9] Y. Zhao and Y. Ban, "GOES-R Time Series for Early Detection of Wildfires with Deep GRU-Network," *Remote Sensing*, vol. 14, no. 17, p. 4347, 2022. doi: 10.3390/rs14174347.

[10] M. Jamali et al., "Saliency Based Fire Detection Using Texture and Color Features," in *2020 28th Iranian Conference on Electrical Engineering (ICEE)*, 2019, pp. 1–5.

[11] J. Huang, J. Zhou, H. Yang, Y. Liu, and H. Liu, "A Small-Target Forest Fire Smoke Detection Model Based on Deformable Transformer for End-to-End Object Detection," *Forests*, vol. 14, no. 1, p. 162, 2023. doi: 10.3390/f14010162.

[12] J. Lian et al., "An Improved Fire and Smoke Detection Method Based on YOLOv7," in *2023 32nd International Conference on Computer Communications and Networks (ICCCN)*, 2023, pp. 1–7.

[13] J. Redmon and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[14] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable Convolutional Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 764–773. doi: 10.1109/ICCV.2017.334.

[15] Wang, Ruoxi et al. "DCN V2: Improved Deep Cross Network and Practical Lessons for Web-scale Learning to Rank Systems." Proceedings of the Web Conference 2021 (2020): n. pag.

[16] Li, Honghao et al. "DCNv3: Towards Next Generation Deep Cross Network for CTR Prediction." (2024).

[17] P. Foggia, A. Saggese and M. Vento, "Real-Time Fire Detection for Video-Surveillance Applications Using a Combination of Experts Based on Color, Shape, and Motion," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 9, pp. 1545-1556, Sept. 2015, doi: 10.1109/TCSVT.2015.2392531.
[18] Songxi Yang, Qunying Huang, Manzhu Yu,Advancements in remote sensing for active fire detection: A review of datasets and methods,Science of The Total Environment,Volume 943,2024,173273,ISSN 0048-9697,https://doi.org/10.1016/j.scitotenv.2024.173273.
[19] Khanam, Rahima and Muhammad Hussain. "YOLOv11: An Overview of the Key Architectural Enhancements." ArXiv abs/2410.17725 (2024): n. pag.
[20] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features for Image Classification," in IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.
[21] Chen B, Wei M, Liu J, et al. EFS-YOLO: a lightweight network based on steel strip surface defect detection[J]. Measurement Science and Technology, 2024, 35(11): 116003.
[22] Hao S, Li X, Peng W, et al. YOLO-CXR: A novel detection network for locating multiple small lesions in chest X-ray images[J]. IEEE Access, 2024.
[23] Tan, Mingxing et al. "EfficientDet: Scalable and Efficient Object Detection." 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019): 10778-10787.
[24] Kandasamy Gajendiran, Sabariswaran Kandasamy, Mathiyazhagan Narayanan,Influences of wildfire on the forest ecosystem and climate change: A comprehensive study,Environmental Research,Volume 240, Part 2,2024,117537,ISSN 0013-9351,https://doi.org/10.1016/j.envres.2023.117537.
[25] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, and S. W. Baik, "Efficient Deep CNN-Based Fire Detection and Localization in Video Surveillance Applications," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 49, no. 7, pp. 1419-1434, July 2019.
[26] S. Frizzi, R. Kaabi, M. Bouchouicha, J. M. Ginoux, E. Moreau, and F. Fnaiech, "Convolutional neural network for video fire and smoke detection," in Proc. IECON 42nd Annual Conference of the IEEE Industrial Electronics Society, 2016, pp. 877-882.
[27] W. S. Mseddi, R. Ghali, M. Jmal and R. Attia, "Fire Detection and Segmentation using YOLOv5 and U-NET," 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 2021, pp. 741-745.
[28] C. Lu, M. Lu, X. Lu, M. Cai, and X. Feng, "Forest Fire Smoke Recognition Based on Multiple Feature Fusion," in IOP Conf. Ser.: Mater. Sci. Eng., vol. 435, 2018, p. 012006. doi: 10.1088/1757-899X/435/1/012006.
[29] Cao, X., Wu, J., Chen, J. et al. Complex Scenes Fire Object Detection Based on Feature Fusion and Channel Attention. Arab J Sci Eng (2024). https://doi.org/10.1007/s13369-024-09471-y
[30] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.
[31] G. Jocher et al., "Ultralytics YOLOv5," 2021. [Online]. Available: https://github.com/ultralytics/yolov5.
[32] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable ConvNets v2: More Deformable, Better Results," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9308-9316.
[33] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, pp. 12993-13000, 2020.
[34] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 658-666.
[35] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," Information, vol. 11, no. 2, p. 125, 2020.
[36] S. Gao, M. Cheng, K. Zhao, X. Zhang, M. Yang, and P. Torr, "Res2Net: A New Multi-scale Backbone Architecture," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 2, pp. 652-662, 2021.
[37] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2988.
[38] K. Liu and G. Mattyus, "Fast Multiclass Vehicle Detection on Aerial Images," in IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 9, pp. 1938-1942, Sept. 2015,.
[39] Z. Zhou, Y. Shi, Z. Gao, and S. Li, "Wildfire smoke detection based on local extremal region segmentation and surveillance," Fire Safety Journal, vol. 85, pp. 50-58, 2016.
[40] C. Wang, A. Bochkovskiy, and H. Liao, "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 (CVPR), 2023, pp. 7464-7475.

ZHIPENG XUE **Zhipeng Xue** is currently pursuing a Master's degree in Electronic Information Engineering at the School of Electronic and Information, Xijing University, Xi'an, China. He received the B.S. degree in Intelligent Science and Technology from Shanxi Institute of Technology, Yangquan, China, in 2024.

Since 2024, he has been studying at Xijing University, focusing on advanced topics in computer vision and large-scale models. His research interests include deep learning, generative models, and their applications in computer vision. He has participated in several research projects, gaining valuable practical experience and focusing on applying theoretical knowledge to solve real-world problems.

During his undergraduate studies, he was involved in multiple projects related to intelligent systems and machine learning, which laid a solid foundation for his current research. He is particularly interested in the development of efficient algorithms for object detection and image recognition, with a focus on improving the accuracy and robustness of these systems.

LINGYUN KONG **Lingyun Kong** is a distinguished professor and the head of the Control Engineering discipline at Xijing University, as well as a supervisor for master's students. He holds a postdoctoral degree in Engineering. His primary research interests encompass robotics, artificial intelligence, and nonlinear control theory and applications.

Professor Kong has received numerous accolades for his contributions to academia and education. In 2015, he was honored with the title of Provincial Academic and Technical Leader. In 2016, he was recognized as an Outstanding Educational Manager at the provincial level. Additionally, in 2020, he was awarded the prestigious Shaanxi Province Teacher Ethics Model title.

Over his career, Professor Kong has dedicated himself to cutting-edge research in robotics control technology, intelligent perception, natural language processing, and nonlinear chaotic dynamics. He has published over 50 papers in both domestic and international journals, including more than 10 SCI-indexed articles. He has successfully led the completion of 12 provincial and municipal-level scientific research projects and 8 enterprise-sponsored projects. Furthermore, he has been acknowledged as an Outstanding Technology Talent in private education at the municipal level.

HAIYANG WU **Haiyang Wu** is currently pursuing a Master's degree in Electronic Information Engineering at the School of Electronic and Information, Xijing University, Xi'an, China. He received the B.S. degree in Intelligent Science and Technology from Shanxi Institute of Technology, Yangquan, China, in 2024.

Since 2024, he has been studying at Xijing University, focusing on advanced topics in computer vision and large-scale models. His research interests include deep learning, generative models, and their applications in computer vision. He has participated in several research projects, gaining valuable practical experience and focusing on applying theoretical knowledge to solve real-world problems.

During his undergraduate studies, he was involved in multiple projects related to intelligent systems and machine learning, which laid a solid foundation for his current research. He is particularly interested in the development of efficient algorithms for object detection and image recognition, with a focus on improving the accuracy and robustness of these systems.

JIALE CHEN **Jiale Chen** is currently pursuing a Master's degree in Electronic Information Engineering at the School of Electronic and Information, Xijing University, Xi'an, China. He received the B.S. degree in Automation from Yuanchai College, Shaoxing University, Shaoxing, China, in 2024.

Since 2024, he has been studying at Xijing University, focusing on advanced topics in ROS (Robot Operating System), robotic control systems, and computer vision. His research interests include embedded deep learning, the application of robotic control algorithms within the ROS framework, and computer vision for robotic perception. He has participated in several research projects, gaining valuable practical experience and focusing on applying theoretical knowledge to the control and optimization of real-world robotic systems.

During his undergraduate studies, he was involved in multiple projects related to ROS (Robot Operating System), robotic control systems, and computer vision. These projects laid a solid foundation for his current research. He is particularly interested in optimizing robotic control algorithms, with a focus on enhancing the real-time performance and stability of these systems. Additionally, he explores the integration of computer vision techniques to improve robotic perception and decision-making capabilities, aiming to develop more intelligent and autonomous robotic systems.