


## Article

# ESFD-YOLOv8n: Early Smoke and Fire Detection Method Based on an Improved YOLOv8n Model

Dilshodjon Mamadaliev <sup>1</sup>, Philippe Lyonel Mbouembe Touko <sup>1</sup>, Jae-Ho Kim <sup>1,2</sup> and Suk-Chan Kim <sup>1,\*</sup>

<sup>1</sup> Department of Electronics Engineering, Pusan National University, Busan 46241, Republic of Korea; mamadalievdilshodjon@gmail.com (D.M.); lyoneltouko@gmail.com (P.L.M.T.); jhkim@pusan.ac.kr (J.-H.K.)

<sup>2</sup> Exsolit Research Center, Yangsan 50611, Republic of Korea

\* Correspondence: sckim@pusan.ac.kr

**Abstract:** Ensuring fire safety is essential to protect life and property, but modern infrastructure and complex settings require advanced fire detection methods. Traditional object detection systems, often reliant on manual feature extraction, may fall short, and while deep learning approaches are powerful, they can be computationally intensive, especially for real-time applications. This paper proposes a novel smoke and fire detection method based on the YOLOv8n model with several key architectural modifications. The standard Complete-IOU (CIoU) box loss function is replaced with the more robust Wise-IOU version 3 (WIoUv3), enhancing predictions through its attention mechanism and dynamic focusing. The model is streamlined by replacing the C2f module with a residual block, enabling targeted feature extraction, accelerating training and inference, and reducing overfitting. Integrating generalized efficient layer aggregation network (GELAN) blocks with C2f modules in the neck of the YOLOv8n model further enhances smoke and fire detection, optimizing gradient paths for efficient learning and high performance. Transfer learning is also applied to enhance robustness. Experiments confirmed the excellent performance of ESFD-YOLOv8n, outperforming the original YOLOv8n by 2%, 2.3%, and 2.7%, with a mean average precision (mAP@0.5) of 79.4%, precision of 80.1%, and recall of 72.7%. Despite its increased complexity, the model outperforms several state-of-the-art algorithms and meets the requirements for real-time fire and smoke detection.

**Keywords:** early smoke and fire detection; YOLOv8n; residual block; GELAN; WIoUv3; artificial intelligence



**Citation:** Mamadaliev, D.; Touko, P.L.M.; Kim, J.-H.; Kim, S.-C. ESFD-YOLOv8n: Early Smoke and Fire Detection Method Based on an Improved YOLOv8n Model. *Fire* **2024**, *7*, 303. <https://doi.org/10.3390/fire7090303>

Academic Editors: Alistair M. S. Smith, António Vieira, Washington Rocha and Marcos Francos

Received: 29 June 2024

Revised: 9 August 2024

Accepted: 23 August 2024

Published: 27 August 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

Despite fire being an essential tool for human development, providing warmth and light, and enabling various technological advancements, it can quickly transform into a destructive force when it escapes our control. Recent statistics underscore the devastating impact of uncontrolled fires on both human lives and infrastructure. In 2022, the United States experienced an estimated 1.5 million fires, resulting in thousands of fatalities, injuries, and USD billions in property damage [1]. The World Health Organization reported over 180,000 annual deaths globally caused by burns and fire-related incidents, many of which occur in densely populated areas [2]. These alarming figures highlight the critical need for advanced fire detection systems capable of the early and accurate identification of fire incidents, potentially saving countless lives and minimizing property loss.

Implementing advanced smoke and fire detection systems significantly improves safety and resource management by accelerating response times and minimizing false alarms. Fire detection has long relied on traditional sensor-based methods, including heat, smoke, flame, and gas sensors. These systems, such as those described in [3–7], are used widely because of their effectiveness in various environments. On the other hand, these sensors often face significant drawbacks, particularly regarding response time and false alarm rates [8–14]. For example, heat sensors may react too slowly in fast-spreading

fires, while smoke sensors can be triggered by non-fire-related particulates, leading to frequent false alarms. Moreover, flame detectors, which rely on the optical spectrum, can be ineffective in smoky conditions where visibility is low. Gas sensors, however, require specific combustible gases to be present and may not detect all types of fires efficiently.

To overcome these limitations, researchers have explored machine learning (ML) methods that analyze color, texture, and motion in visual data for fire detection. Previous studies [15–18] used color-based models to identify fire pixels, leveraging the characteristic color features of flames. On the other hand, these models often struggle with varying lighting conditions and similar-colored objects. Texture-based approaches examine the spatial patterns of fire but can be confounded by complex background textures, as discussed in [19,20]. Motion-based detection, which has been used in previous studies, capitalizes on the dynamic nature of fire, distinguishing it from static objects [21–23]. Nevertheless, these methods are challenged by other moving objects, such as humans or vehicles, which can result in false positives. Despite these advances, ML methods still require robust preprocessing and feature extraction, which can be computationally intensive and less effective in real-time applications [24–26].

The advent of deep learning (DL) has brought significant advances in fire detection technology, addressing many of the challenges faced by traditional and ML methods. In particular, convolutional neural networks (CNNs) have shown remarkable success because of their ability to extract hierarchical features automatically from raw input data. For example, the study by Saydirasulovich et al. [27] achieved an average precision (AP) of 79.4% in identifying fires and smoke within intricate scenarios. In addition, the FireNet architecture [28] is lightweight enough to be deployable on embedded platforms, such as Raspberry Pi, balancing accuracy and resource efficiency.

Efforts to boost the speed and real-time performance of fire detection algorithms have yielded promising results. Adaptations of YOLOv5 [29–33] and the MobileNetV2-SSD model [34–36] enable real-time fire detection on embedded systems and video surveillance applications, achieving a harmony between speed and accuracy. Furthermore, recent research has focused on developing fire detection systems capable of operating in diverse environments. A previous study [37] proposed a hybrid system combining local binary pattern convolutional neural network (LBP-CNN) and YOLOv5 architectures. They reported a fire and smoke detection precision rate of 96.25% for a normal scenario, 93.2% for a foggy environment, and a combined detection average precision rate of 94.59% across varied settings. This approach highlights the importance of using spatial information for accurate detection in diverse environments. The authors of [38] incorporated Ghost modules [39] and attention mechanisms, resulting in a 27% parameter reduction, a 2.9% increase in mAP<sub>50small</sub>, and an FPS of 24.4. In addition, incorporating temporal information alongside spatial data has shown promising results. Studies, such as de Venâncio [40], reported how combining both aspects helps to reduce false positives without compromising accuracy or processing time, proving advantageous over common approaches. These novel methodologies pave the way for more robust and adaptable fire detection systems.

Similarly, research has delved into wildfire detection across various fire and smoke images, tailoring deep learning architectures to differing lighting and weather conditions [41–47]. Other investigations have explored the use of transformer-based models, which offer improved feature extraction and classification capabilities for complex fire and smoke patterns [48–50]. Additionally, advancements have been made in leveraging multi-spectral and hyperspectral imaging combined with machine learning techniques to enhance the detection accuracy and responsiveness of wildfire detection systems [51–53].

Dedicated efforts have also been directed towards smoke detection, a fundamental piece of early fire warning systems. By incorporating techniques, such as the channel prior dilatation attention module and mixed-classification detection head, advanced object detection architectures inspired by the YOLOv8 model have demonstrated promising results in enhancing smoke detection capabilities for static images and video sequences [54]. Furthermore, the application of generative adversarial networks (GANs) has shown potential in

augmenting training datasets, thus improving the robustness of smoke detection algorithms in varied conditions [55–58]. Research has also investigated the role of edge computing in smoke detection, emphasizing the benefits of distributed processing and real-time analytics in minimizing detection latency and improving system scalability [59–62].

Although AI-driven smoke and fire detection has seen significant advancements, achieving a balance between high accuracy and resource efficiency remains challenging. This paper addresses this issue by introducing a novel early smoke and fire detection system based on YOLOv8n [63], which was designed for broad applicability across diverse environments. The unique advantages of the system are highlighted through a comparative analysis with existing fire detection models, including those within the well-established YOLO family [64], all trained on the same dataset.

The contributions of this study include the following:

- The backbone network was enhanced by substituting C2f blocks with residual blocks [65]. This architectural change improves feature extraction and information flow, producing more robust representations of smoke and fire patterns;
- C2f-integrated generalized efficient layer aggregation network (GELAN [66]) blocks (C2fGELAN) were introduced in the neck network, replacing standard C2f blocks. C2fGELAN is a powerful and state-of-the-art feature enhancement block that refines feature maps, enabling the model to better distinguish among smoke, fire, and background elements;
- The standard Complete Intersection over Union (CIoU [67]) loss function was replaced with the Wise Intersection over Union version 3 (WIoUv3) [68] loss during training. The accuracy of the bounding boxes delineating smoke and fire was significantly improved, resulting in a more precise localization of these hazards;
- Transfer learning was applied, further enhancing the model's robustness and ability to generalize across various environments and conditions;
- Extensive experiments, including comparisons with the YOLO series and other methods, showed that this model achieved the best overall accuracy and efficiency in detecting smoke and fire in diverse scenarios.

The remainder of this work is structured as follows. Section 2 details the materials and methods, including a review of relevant literature and the model architecture. Section 3 outlines the datasets, evaluation metrics, and experimental setup, followed by a presentation and analysis of the results, highlighting the strengths and weaknesses of this approach. Section 4 concludes this paper by summarizing the key findings, discussing the limitations identified through error analysis, and suggesting potential directions for future research.

## 2. Materials and Methods

### 2.1. YOLO Algorithm

The “you only look once” (YOLO) series revolutionized real-time object detection with speed, accuracy, and adaptability. Since 2015, each iteration has improved upon its predecessors. YOLO models allow real-time processing using a single-stage architecture. Despite prioritizing speed, they show competitive accuracy.

The YOLO series has seen significant advances over the years. YOLOv1 introduced the groundbreaking concept of real-time object detection with a single-stage network. YOLOv2, also known as YOLO 9000 [69], showed improved accuracy and speed through techniques such as anchor boxes, multi-scale training, and better backbone networks. YOLOv3 [70] introduced a feature pyramid network (FPN) [71] for the better detection of objects on different scales, further improving accuracy. YOLOv4 [72] incorporated various architectural and training optimizations, increasing speed and accuracy. YOLOv5 [73] focused on ease of use, model size optimization, and training flexibility, making it popular for deployment on diverse platforms. YOLOv6 [74] emerged as an anchor-free version, streamlining the detection process. YOLOv7 [75] focused on performance optimization, incorporating new training, and extending the efficient layer aggregation network (ELAN [76]) to E-ELAN, thereby improving efficiency and prioritizing both accuracy and speed. YOLOv8 integrated

advances in anchor-free detection, loss functions, and training strategies. YOLOv9 introduced the novel concepts of programmable gradient information (PGI) and the GELAN, achieving state-of-the-art results on the MS COCO dataset. The most recent iteration, YOLOv10 [77], developed by researchers at Tsinghua University, eliminated the non-maximum suppression (NMS) step, optimized the architecture for enhanced performance, and reduced computational overhead in end-to-end real-time object detection.

Since its inception, researchers have continuously enhanced the accuracy and robustness of YOLO through various approaches, such as refined architectures, attention mechanisms, and optimized training strategies [78–81]. The adaptability of YOLO has been demonstrated in diverse applications such as aerial imagery [82–85], underwater environments [86–88], agriculture [89,90], and medical imaging [91–93]. Efforts to boost the speed and efficiency of YOLO for real-time use cases have involved techniques like model compression and hardware acceleration [94–96]. Moreover, the scope of YOLO has expanded beyond object detection, integrating with tasks like semantic segmentation and action recognition for a more comprehensive scene understanding [97–100].

## 2.2. YOLOv8n Network

This study focused on YOLOv8n, the nano version of the YOLOv8 object detection model developed by Ultralytics LLC located in Los Angeles, CA, United States. Despite advances in YOLOv9 and YOLOv10, YOLOv8n was selected for its proven effectiveness and balanced approach within the YOLO family. The introduction of PGI by YOLOv9 complicates training, while the elimination of Non-Maximum Suppression (NMS) by YOLOv10 raises resource requirements during training. Consequently, YOLOv8n is more suitable for resource-constrained environments while delivering competitive performance. The architecture of YOLOv8n (Figure 1) consists of the following three main components: a backbone network, a neck network, and a head network.

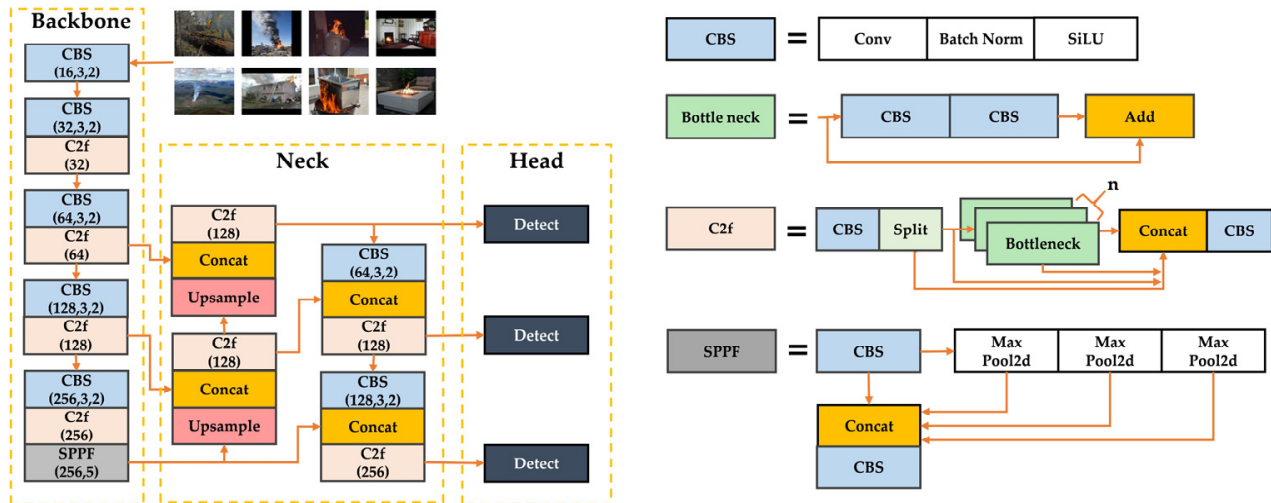
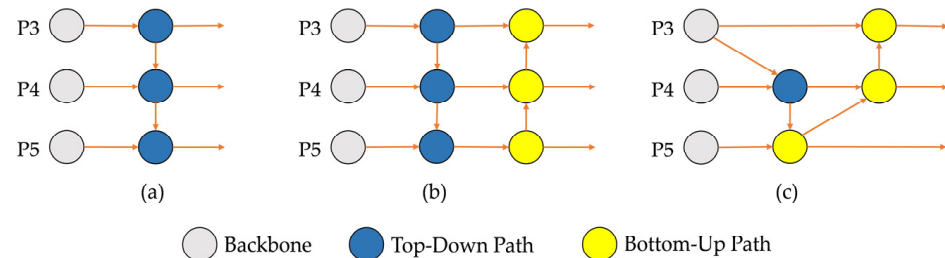


Figure 1. Architecture of YOLOv8n.

YOLOv8n uses a modified version of the CSPDarknet [101] architecture as its backbone. It comprises stacked modules, including CBS (light blue colored blocks in Figure 1) (convolutional layer (Conv) + batch normalization (BatchNorm [102]) + Sigmoid Linear Unit (SiLU [103])) and C2f modules (warm peachy colored blocks in Figure 1), culminating in the spatial pyramid pooling fast (SPPF) module (gray colored blocks in Figure 1). While CBS and C2f modules aid in feature extraction, the SPPF module enhances feature expression using max pooling on previously pooled features, improving module efficiency.

The neck of YOLOv8n features a combination of path aggregation network (PANet [104]) and FPN layers (Figure 2). This hybrid approach enhances the ability of the model to aggregate features from different scales, which is crucial for detecting objects of varying sizes. PANet

improves information flow among layers by adding bottom-up paths, enriching the semantic information at each level. FPN, on the other hand, helps build high-level feature maps with strong semantics. Additionally, the Upsample blocks (light coral colored blocks in Figure 1) enhance feature map resolution to maintain spatial accuracy, while the Concatenation blocks (gold colored blocks in Figure 1) merge features from different layers to integrate both high-resolution and semantically rich information. Together, these components ensure that the features used by the head are rich and diverse, enhancing detection accuracy.



**Figure 2.** Feature network design—(a) FPN, (b) PANet, and (c) FPN + PANet—YOLOv8 neck.

The head of YOLOv8n (charcoal blue colored blocks in Figure 1) predicts the bounding boxes, objectness scores, and class probabilities. It features a fully convolutional structure with multiple output layers, each corresponding to a different scale. This multi-scale prediction approach ensures the model can detect small, medium, and large objects. YOLOv8n uses an anchor-free approach that simplifies the design and reduces computational complexity. This approach predicts the bounding boxes directly without predefined anchor boxes, improving flexibility and performance.

### 2.3. Generalized Efficient Layer Aggregation Network (GELAN)

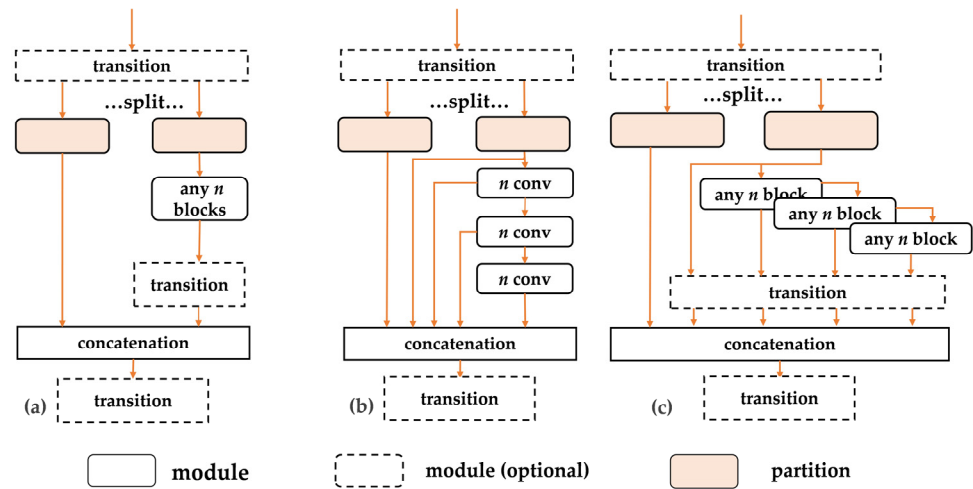
The efficient layer aggregation network (ELAN [76]) has revolutionized object detection tasks. ELAN enhances feature learning and computational efficiency by optimizing gradient paths within deep convolutional networks. Its integration of cross-stage partial connections (CSPs) and stacked computational blocks facilitates improved gradient flow during training, resulting in faster model convergence and excellent performance in identifying smoke and fire in their early stages. Moreover, the design of ELAN mitigates the risk of overfitting, making it adaptable to diverse datasets encompassing various smoke and fire scenarios.

Nevertheless, the focus of ELAN on stacking convolutional layers restricts the flexibility to integrate other computational blocks that might be advantageous for detecting the subtle early signs of smoke and fire. This limitation can hinder its effectiveness in complex real-world scenarios, where varying environmental conditions and potential obstructions can impede accurate early detection.

In YOLOv9, GELAN addresses these limitations (Figure 3). Building upon the strengths of ELAN, GELAN broadens the concept of layer aggregation, enabling the incorporation of any computational block, not just convolutional layers. This enhanced flexibility allows the integration of specialized modules like attention mechanisms or unique feature extractors, which are better equipped to capture the distinctive characteristics of early smoke and fire. GELAN retains the emphasis of ELAN on gradient path optimization, ensuring efficient learning and improved performance.

GELAN offers a more versatile and powerful architecture for object detection tasks because it merges the best aspects of CSPNet and ELAN. Its ability to integrate diverse computational blocks seamlessly while maintaining high efficiency and accuracy makes it an invaluable asset in pursuing enhanced early detection capabilities for critical real-time applications.





**Figure 3.** Architecture of GELAN. (a) CSPNet, (b) ELAN, and (c) GELAN.

#### 2.4. Wise Intersection over Union Version 3 (WIoUv3)

In object detection, selecting the right loss function for bounding box regression is crucial for accuracy. The primary goal is to minimize differences between the predicted boxes of the model and the actual ground truth boxes around objects. This choice is particularly important in challenging applications, such as smoke and fire detection, where training data annotations may be noisy or incomplete.

This study addressed this by replacing the standard CIoU loss function in YOLOv8n with WIoUv3. WIoUv3 offers significant advantages, particularly when dealing with suboptimal training data.

WIoUv3 features a two-layer attention mechanism and a dynamic non-monotonic focusing mechanism. The attention mechanism helps the model focus on relevant features within bounding box predictions. The dynamic focusing mechanism evaluates the quality of candidate boxes (anchors) through outlier analysis, reducing the impact of noisy annotations and preventing the model from being penalized for minor geometric discrepancies. When the predictions are close to the ground truth, WIoUv3 de-emphasizes geometric factors, aiding efficient learning and better generalization. The precise mathematical formulation of this loss function is expressed as Equations (1)–(7).

$$S_u = wh + w_{gt}h_{gt} - W_iH_i, \quad (1)$$

$$IoU = \frac{W_iH_i}{S_u}, \quad (2)$$

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

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{W_g^2 + H_g^2}\right), \quad (4)$$

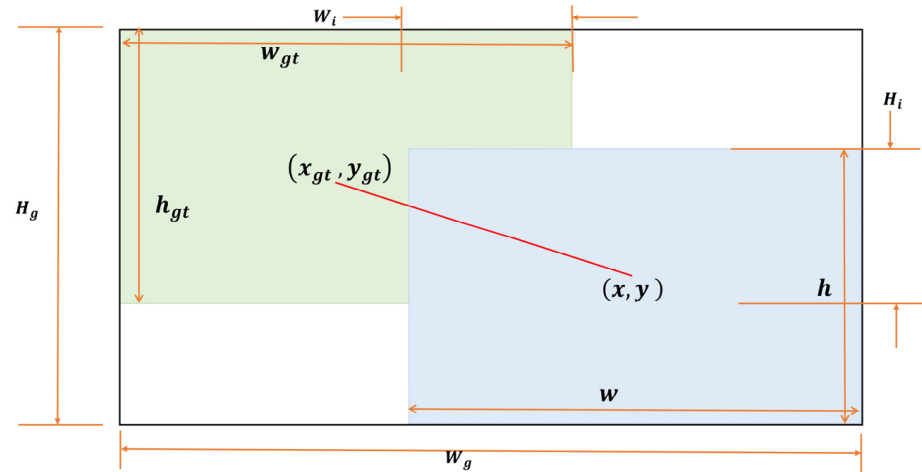
$$L_{WIoUv1} = R_{wIoU}L_{IoU}, \quad (5)$$

$$L_{WIoUv3} = rL_{wIoUv1} \quad (6)$$

$$r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \quad (7)$$

where  $IoU$  is the intersection over union [105];  $r$  is the gradient gain,  $\alpha$  and  $\beta$  are the anomaly levels of the predicted box, with lower values indicating higher quality. Smaller gradient gains are assigned to boxes with larger anomalies to reduce the harmful gradients during training.  $\alpha$  and  $\delta$  are the hyperparameters controlling gradient gain. Figure 4 presents the other parameters, with blue and green boxes representing predicted and ground truth boxes, respectively. Coordinates are denoted by  $(x, y)$  and  $(x_{gt}, y_{gt})$ , while

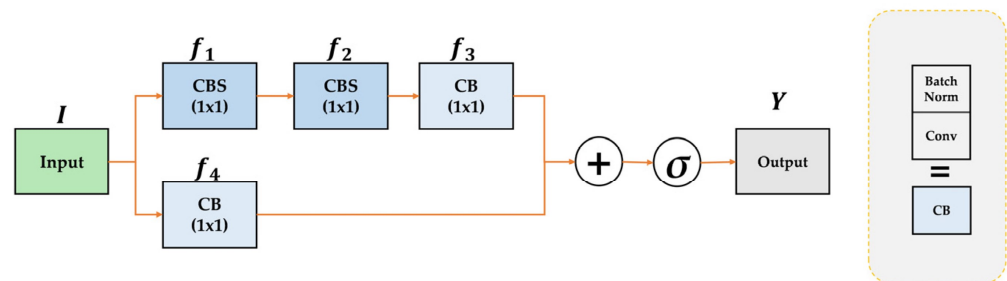
the heights and widths are  $(h, w)$  and  $(h_{gt}, w_{gt})$ .  $H_i$  and  $W_i$  are the height and width of the intersection rectangle, while  $H_g$  and  $W_g$  are the height and width differences between the combined boxes and their intersection.  $S_u$  in Equation (1) is the union area of the predicted box and the ground truth box.



**Figure 4.** Calculation process of WIoUv3.

### 2.5. Residual Block (ResBlock)

Residual blocks (ResBlocks) are integral to deep neural networks, particularly in architectures like ResNet (Figure 5). They incorporate skip connections, enabling them to learn residual functions relative to the layer input rather than attempting to learn functions from scratch.



**Figure 5.** Architecture of a residual block.  $\sigma$  represents the SiLU activation function and  $+$  represents addition.

Therefore, residual blocks aim to fit a residual mapping of the desired underlying mapping. This is achieved by recasting the original mapping into a residual form, denoted as Equation (12).

$$f_1 = CBS_{1 \times 1}(I), \quad (8)$$

$$f_2 = CBS_{3 \times 3}(f_1), \quad (9)$$

$$f_3 = CB_{1 \times 1}(f_2), \quad (10)$$

$$f_4 = CB_{1 \times 1}(I), \quad (11)$$

$$Y = \sigma(f_3 + f_4) \quad (12)$$

where  $I$  represents the input to the block, the CBS block (light blue colored blocks in Figure 5) denotes a sequence of a convolutional layer, batch normalization layer, and SiLU activation function, while the CB block (powder blue colored blocks in Figure 5) consists of a convolutional layer followed by a batch normalization layer, and  $\sigma$  represents the SiLU activation function.

The residual blocks comprise convolutional layers, batch normalization layers, and SiLU activation functions. SiLU is the activation function within these blocks. Batch normalization plays a crucial role in the proper initialization of neural networks, ensuring that activations throughout the network conform to a unit Gaussian distribution at the outset of training.

The rationale behind using residual blocks lies in the ease of optimizing the residual mapping compared with the original unreferenced mapping. Adjusting the residual towards zero is simpler if an identity mapping is optimal rather than attempting to learn the identity mapping directly through a stack of nonlinear layers. The inclusion of skip connections facilitates the learning of identity-like mappings, enhancing the capacity of the network to learn complex functions. Consequently, the residual acts as a correction to the original mapping.

## 2.6. ESFD-YOLOv8n Architecture

This paper introduces ESFD-YOLOv8n (Early Smoke Fire Detection-YOLOv8n), a specialized YOLOv8n model for early and accurate smoke and fire detection (Figure 6). The proposed model refines the following three critical components: feature extraction, the neck network, and box prediction (Table 1). Residual blocks (ResBlock—warm peachy colored blocks in the Backbone in Figure 6) are integrated into the architecture to strengthen the ability of the model to extract meaningful features from images, making it more resilient to variations in lighting, perspective, and other visual complexities often found in challenging environments. This enhancement significantly improves the ability of the model to identify subtle visual cues associated with smoke and fire, even in their earliest stages.

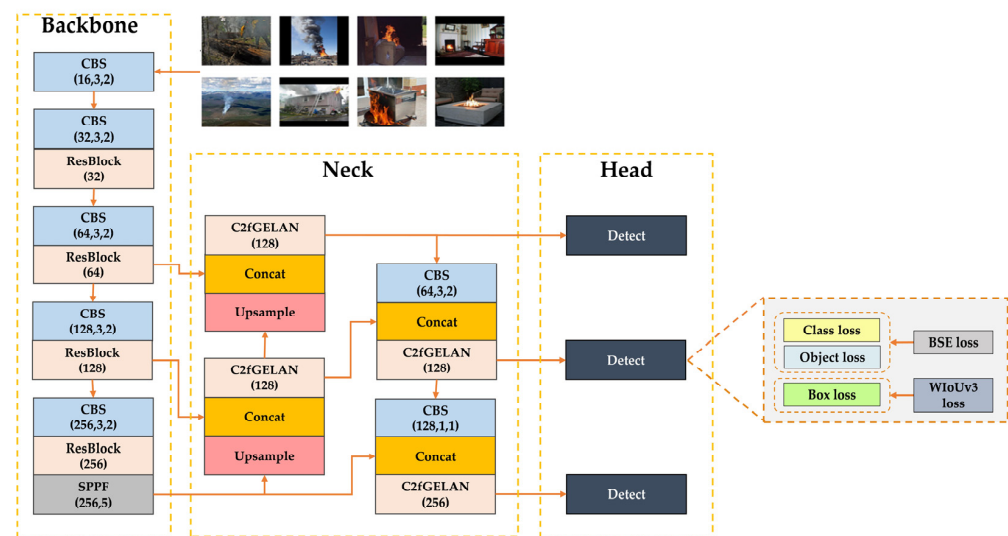


Figure 6. ESFD-YOLOv8n model architecture.

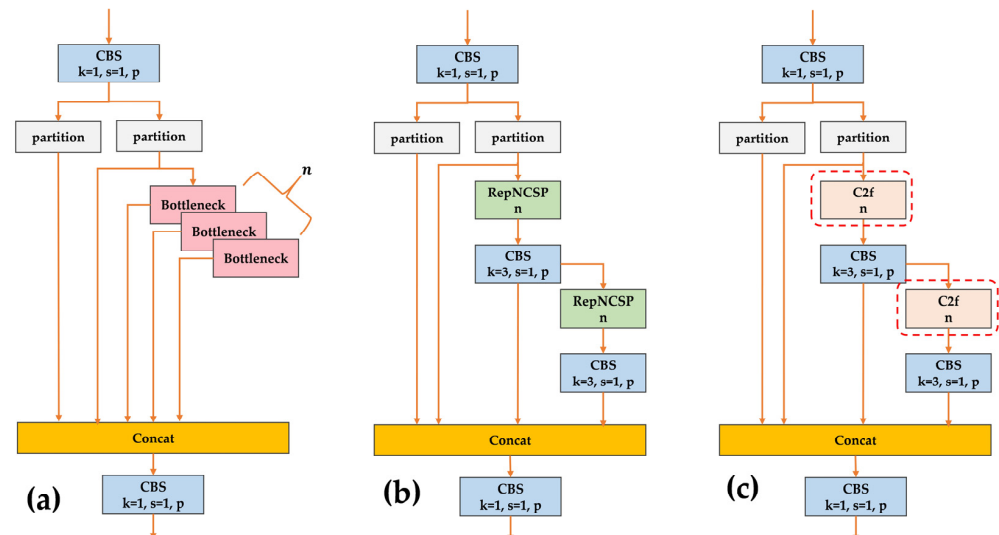
Table 1. Comparison of the proposed and YOLOv8n architectures.

Features	The Proposed Architecture	YOLOv8n
Backbone	ResBlock	C2f
Neck	C2fGELAN	C2f
Bounding box loss function	WIoUv3	CIoU
Techniques used	ResBlock, WIoUv3	-
Novel technique introduced	C2fGELAN	-

In the neck network, this study introduced C2fGELAN—the integration of C2f and GELAN blocks (warm peachy colored blocks in the neck in Figure 6)—enriching extracted features with deeper semantic information (Figure 7). This comprehensive understanding of image content empowers the model to distinguish smoke and fire from other visual



elements more precisely, minimizing false alarms and enhancing overall detection accuracy. The WIoUv3 loss function (light steel blue colored block in the neck in Figure 6) was used as the box predictor loss function, replacing the standard CIoU. This strategic choice significantly enhanced the accuracy of the model in detecting smoke and fire at the earliest stages. The WIoUv3 loss function provided a more refined and precise approach, allowing the model to identify potential fire hazards more accurately.



**Figure 7.** C2fGELAN architecture. (a) C2f, (b) RepNCSP (Reparametrized Cross-Stage Partial and ELAN)—an advanced version of CSP-ELAN used in YOLOv9—and (c) C2fGELAN.

### 3. Experiments and Analysis

#### 3.1. Dataset

The D-Fire dataset [106] was used to train the YOLOv8n-based model. D-Fire is a freely available community resource specifically curated for fire and smoke classification and detection tasks. The strength of the dataset lies in its diversity, containing 21,527 images encompassing a broad spectrum of scenarios. These include indoor and outdoor environments, fires of varying sizes, scenes with varying light conditions, and, importantly, normal scenes devoid of fire or smoke, which often cause false alarms in conventional algorithms. In particular, the dataset also includes images featuring objects or environments humans commonly misinterpret as fire or smoke, strengthening the ability of the model to navigate the complexities of smoke or fire detection. Figure 8 shows the balanced representation of fire/smoke and normal scenes in D-Fire, which facilitates training a model that can effectively generalize to real-world applications. Table 2 provides a detailed description of the dataset.

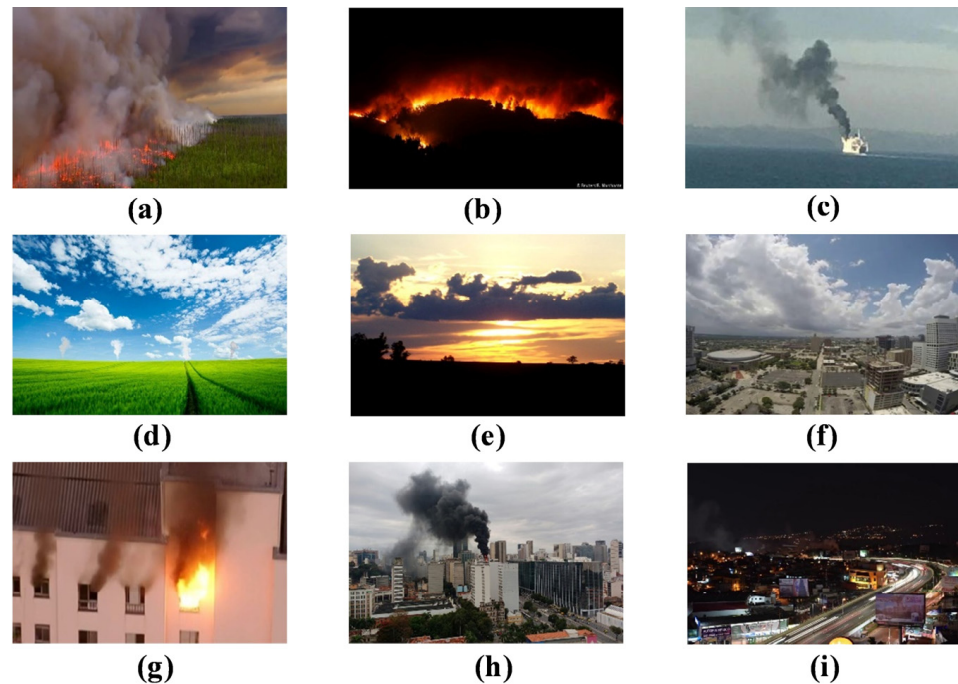
**Table 2.** D-Fire dataset description.

Category	Images
Fire only	1164
Smoke only	5867
Fire and smoke	4658
Background	9838

The D-Fire dataset was divided into 70% training, 20% validation, and 10% testing sets for our early smoke and fire detection research. This ensured sufficient training data, hyperparameter tuning with the validation set, and a separate testing set for evaluating real-world performance. Table 3 lists the image distribution.

**Table 3.** Details of the dataset.

Sets	Images	Percentage
Training	15,069	70
Validating	4305	20
Testing	2153	10



**Figure 8.** Sample images captured in various natural environments within the dataset. (a) Forest fire; (b) forest fire at night; (c) fire and smoke in a ship in the sea; (d) synthetic smoke; (e) fire-like sunset image; (f) smoke-like cloudy sky image; (g) fire and smoke in an apartment; (h) fire and smoke in a city; and (i) fire and smoke at night in a city.

### 3.2. Evaluation Metrics

Performance metrics [107,108] are essential for evaluating the effectiveness of object detection algorithms like YOLOv8. These metrics provide insights into how well the model detects and localizes objects within images. Precision describes the percentage of detected objects that are genuinely the desired class (Equation (13)), while recall indicates the percentage of actual objects the model detects (Equation (14)). An ideal model would have both precision and recall approaching 100%.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}, \quad (13)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (14)$$

These metrics often trade off against each other. A model may achieve high precision by making few predictions, but it might miss many objects, resulting in low recall. Average precision (AP) is used to balance this trade-off (Equation (15)). AP summarizes the precision–recall trade-off of the model across different confidence thresholds.

$$AP = \int_0^1 P(R) dR \quad (15)$$

The mean average precision ( $mAP$ ) provides a single number summarizing performance across all object classes (Equation (16)). The  $mAP$  is the mean of  $AP$  values for each class. The  $mAP$  at a 50%  $IoU$  threshold ( $mAP@50$ ) is a specific instance of the  $mAP$  where the  $IoU$  threshold is set to 0.5. This means a prediction is considered correct if it overlaps with the ground truth by at least 50%.

$$mAP = \frac{1}{N} \sum_{i=1}^N \int_0^1 P_1(R) dR \quad (16)$$

The inference time measures how long the model takes to process an image and generate predictions (Equation (17)). This metric is crucial for real-time applications, directly affecting the responsiveness of the system.

$$Inference\ Time = \frac{Total\ Processing\ Time}{Number\ of\ Images} \quad (17)$$

### 3.3. Configuration Parameters

The proposed method in Anaconda 24.1.2 was constructed and examined using Python distribution version 3.11.7 on a computer with a 12th Gen Intel(R) Core (TM) i9-12900K (24 CPUs), ~3.2 GHz, manufactured by Intel Corporation, located in Santa Clara, United States, 64 GB RAM, and NVIDIA® GeForce RTX™ 4060 Ti 8 GB graphics card, manufactured by NVIDIA Corporation, located in Santa Clara, United States. The model was trained with a batch size of 32 for 300 epochs, starting with an initial learning rate of 0.01. Table 4 lists the configuration parameters for the model.

**Table 4.** Model training configuration parameters.

Parameters	Values
Epochs	300
Batch size	32
Image size	640
Learning rate	0.01
Patience	50
Pretrained	yolov8n.pt

### 3.4. Ablation Study

The contribution of each proposed improvement strategy was evaluated by conducting a series of ablation studies using the D-Fire dataset on our baseline model. The results (Table 5) provide valuable insights into the efficacy of each enhancement.

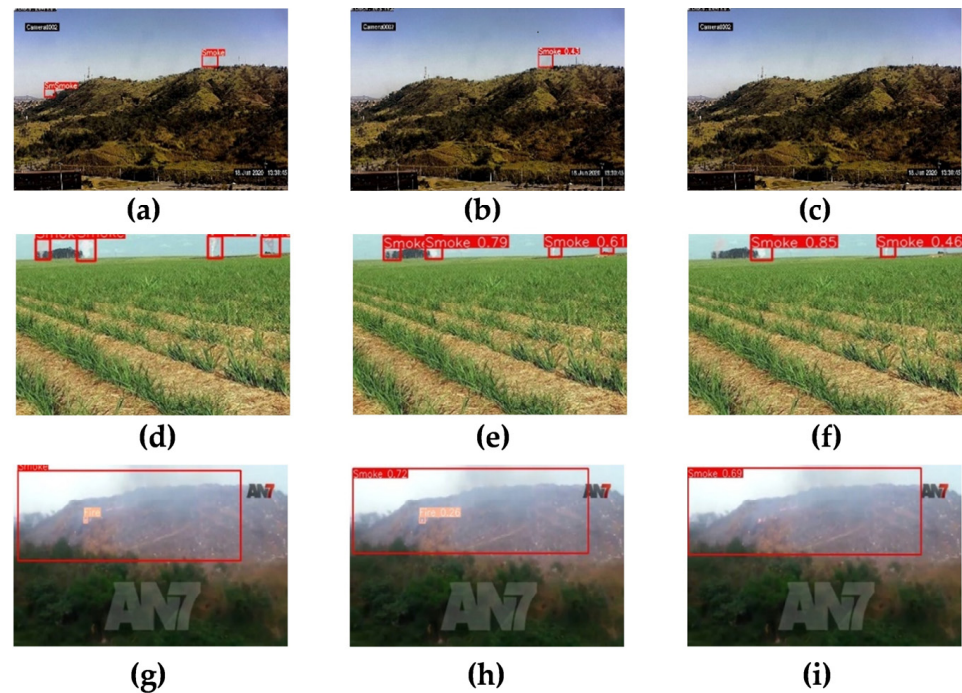
**Table 5.** Detection results after introducing different improvement strategies (the bold data in the table indicate the best results and the check mark (✓) indicates that the method in the corresponding column is applied in the corresponding row).

Baseline	WIoUv3	Residual Block	C2f GELAN	Transfer Learning	P (%)	R (%)	mAP@0.5 (%)	Time (ms)	FLOPs (B)
YOLOv8n	-	-	-	-	77.8	70.0	77.4	0.8	8.1
	✓	-	-	-	79.1	70.6	78.0	<b>0.8</b>	<b>8.1</b>
	✓	✓	-	-	80.8	71.0	78.8	0.9	10.6
	✓	✓	✓	-	<b>81.2</b>	71.2	79.3	1.0	10.8
	✓	✓	✓	✓	80.1	<b>72.7</b>	<b>79.4</b>	1.0	10.8

Table 5 lists the incremental improvements in detection performance achieved by each strategy. In particular, integrating WIoUv3 into the prediction box regression loss enhanced localization significantly, contributing to a 0.6% improvement in mAP@0.5. Replacing the C2f module with residual blocks in the backbone network further refined the attention of

the model to crucial features, boosting mAP@0.5 by an additional 0.8%. The substitution of C2f layers with the C2fGELAN module in the neck resulted in a 0.5% improvement in mAP@0.5, demonstrating the effectiveness of this module in balancing speed and accuracy while improving feature fusion. Finally, applying transfer learning led to a further 0.1% gain in mAP@0.5, highlighting the potential of this technique for model optimization.

The proposed network boosted the detection performance, with a 2.0% average increase in accuracy and notable gains across other metrics, showing that the slight increase in the model size and parameters is a worthwhile trade-off for the substantial accuracy improvement. Furthermore, the model demonstrates proficiency in detecting small-scale objects (Figure 9).



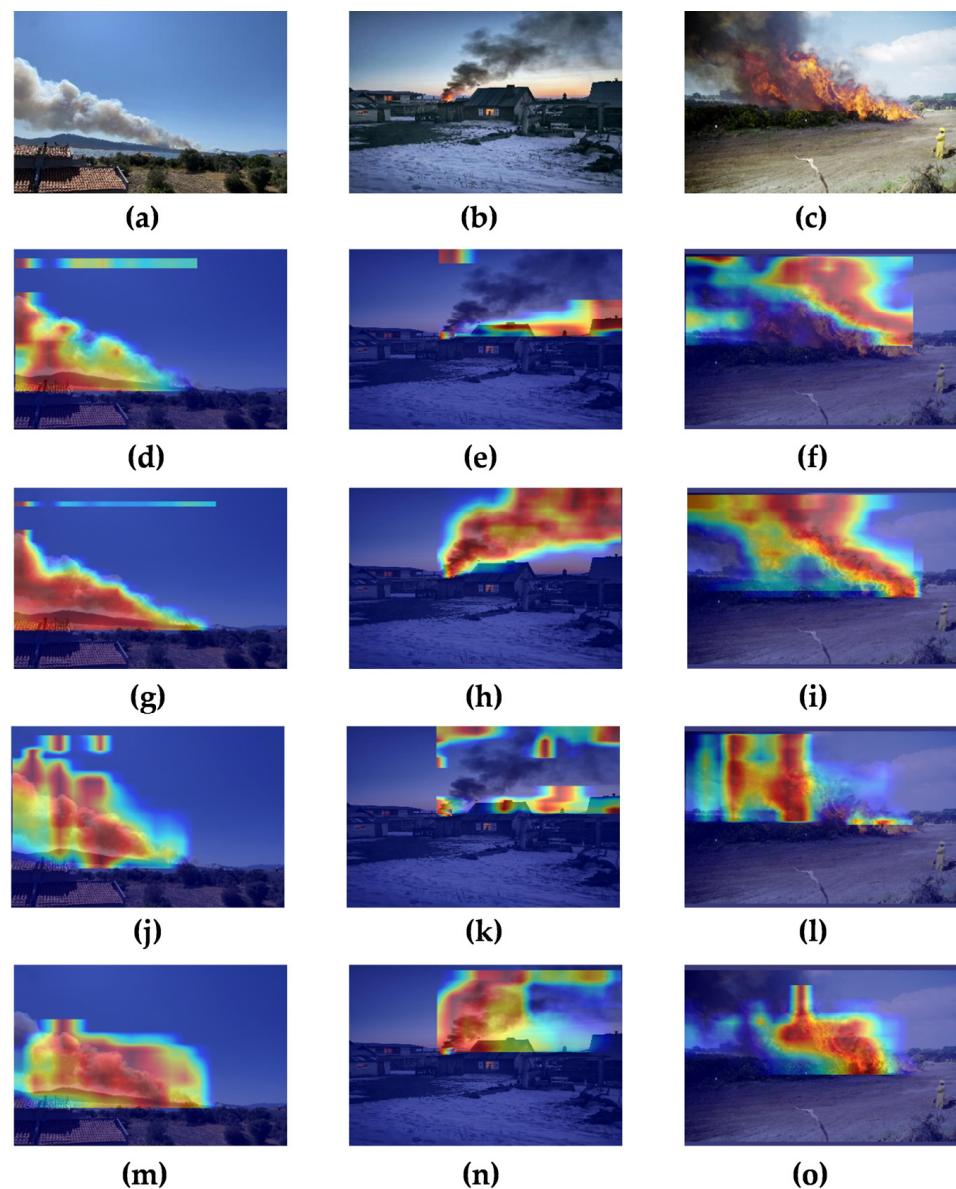
**Figure 9.** Comparison of the detection results of the original and ESFD-YOLOv8n models for small-scaled images. (a,d,g) Labeled images, (b,e,h) images predicted by ESFD-YOLOv8n, and (c,f,i) images predicted by the original YOLOv8n.

### 3.5. Feature Map Visualization

The feature extraction capabilities of the baseline model were enhanced by replacing the existing C2f modules in the backbone with residual blocks and further refining the neck network by substituting C2f layers with C2fGELAN blocks. The impact of these changes was visually assessed using the Eigen-CAM [109] method, focusing on the 8th layer (backbone) and the 21st layer (neck).

Figure 10 provides a clear visualization of the improvements by comparing the proposed model and the original YOLOv8n model on the same images. The bright red and yellow areas on the feature map indicate regions where both models detected significant features, while the blue areas show regions of lower attention. The input images (Figure 10a–c) are presented with enhanced smoke and fire visibility for clarity. The differences between the C2f and ResBlock modules are shown in Figures 10d–f and 10g–i, respectively. Figure 10g–i show the superior ability of the ResBlock module to capture finer details, guiding the backbone network toward more pertinent information. The original and modified necks of YOLOv8n and ESFD-YOLOv8n are depicted in Figures 10j–l and 10m–o, respectively. The enhanced feature maps in Figure 10m–o show the effectiveness of the C2fGELAN upgrade, resulting in heightened resolution and facilitating efficient context aggregation and fusion within the network.





**Figure 10.** Comparison of the feature maps for the baseline vs. proposed methods. (a–c) Input images, (d–f) 8th C2f layer of YOLOv8n, (g–i) 8th ResBlock layer of the proposed model, (j–l) 21st C2f layer of YOLOv8n, and (m–o) 21st C2fGELAN layer of the proposed model.

These visual results confirm that the ESFD-YOLOv8n model outperformed the original model in terms of accuracy, robustness, and efficiency.

### 3.6. Comparison of the ESFD-YOLOv8n with Other Detection Algorithms

The effectiveness of the proposed ESFD-YOLOv8n model was examined through an extensive evaluation of the D-Fire dataset, comparing its performance against a diverse selection of state-of-the-art smoke and fire detection approaches. These included six YOLO series variations (YOLOv5n, YOLOv5s, YOLOX tiny [110], YOLOv6n, YOLOv7 tiny, YOLOv8n, GELAN, YOLOv9, and YOLOv10n) and six other notable approaches [111–116]. The comprehensive results, highlighted in Table 6 with the optimal values in bold, demonstrated the superior performance of ESFD-YOLOv8n. The proposed model achieved a precision of 80.1%, a recall of 72.7%, a mAP@0.5 of 79.4%, and a remarkable inference time of 1.0 ms, highlighting its effectiveness and efficiency in object detection on the D-Fire dataset.

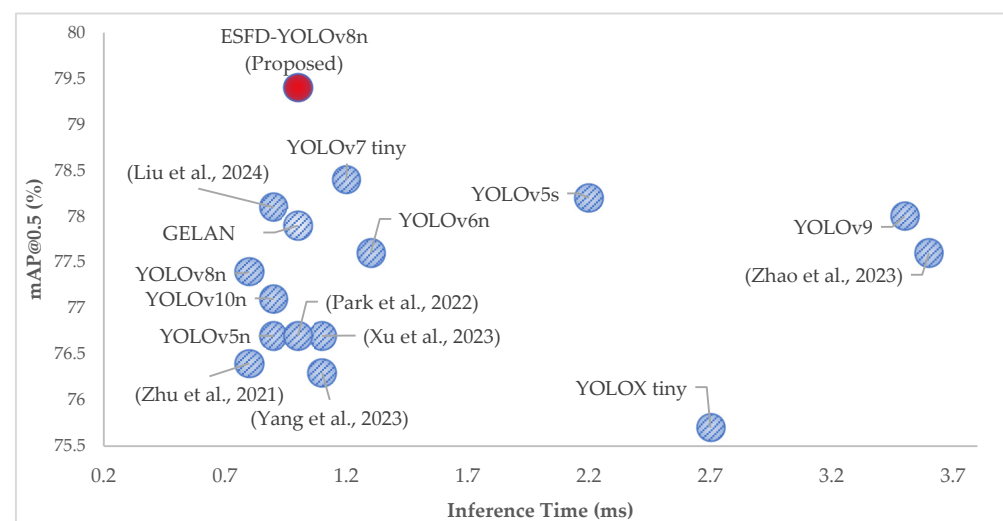


**Table 6.** Comparison of the different models (Bold data in the table indicate the best results).

Methods	Precision (%)	Recall (%)	mAP@0.5 (%)	Time (ms)	FLOPs (B)
Park et al. [111]	78.4	70.3	76.7	1.0	4.2
Yang et al. [112]	79.1	69.7	76.3	1.1	<b>2.3</b>
Xu et al. [113]	78.2	69.9	76.7	1.1	4.5
Liu et al. [114]	79.5	70.8	78.1	0.9	8.2
Zhu et al. [115] *	75.8	71.3	76.4	0.8	7.7
Zhao et al. [116] *	78.0	71.3	77.6	3.6	17.9
YOLOv5n	78.3	69.7	76.7	0.9	4.1
YOLOv5s	79.7	72.7	78.2	2.2	15.8
YOLOX tiny	76.8	69.8	75.7	2.7	15.4
YOLOv6n	77.8	69.8	77.6	1.3	11.3
YOLOv7 tiny	78.3	72.2	78.4	1.2	13.0
YOLOv8n	77.8	70.0	77.4	<b>0.8</b>	8.1
GELAN *	78.4	69.5	77.9	1.0	8.1
YOLOv9 *	78.8	70.3	78.0	3.5	18.1
YOLOv10n	78.3	69.5	77.1	0.9	8.2
ESFD-YOLOv8n (Proposed)	<b>80.1</b>	<b>72.7</b>	<b>79.4</b>	1.0	10.8

\* To match YOLOv8n, the model's size was reduced.

Table 6 presents the exceptional detection performance of the proposed ESFD-YOLOv8n model across all accuracy-related metrics. The proposed ESFD-YOLOv8n model outperformed the other methods, showing higher precision, recall, and mAP@0.5 values (Figure 11).

**Figure 11.** Comparison with different YOLO series and methods [111–116]. Speed versus accuracy (mAP@0.5).

In particular, ESFD-YOLOv8n achieves notable improvements in precision over the existing methods, exceeding [111–116], YOLOv5n, YOLOv5s, YOLOX tiny, YOLOv6n, YOLOv7 tiny, YOLOv8n, GELAN, YOLOv9n, and YOLOv10n by 1.7%, 1.0%, 1.9%, 0.6%, 4.3%, 2.1%, 1.8%, 0.4%, 3.3%, 2.3%, 1.8%, 2.3%, 1.7%, 1.3%, and 1.8%, respectively. Furthermore, it outperformed the [111–114], YOLOv5n, YOLOX tiny, YOLOv6n, YOLOv7 tiny, YOLOv8n, GELAN, YOLOv9, and YOLOv10n methods in recall by 2.4%, 3.0%, 2.8%, 1.9%, 1.4%, 1.4%, 3.0%, 2.9%, 2.9%, 0.5%, 2.7%, 3.2%, 2.4%, and 3.2%, respectively, achieving results comparable to YOLOv5s. Similarly, ESFD-YOLOv8n demonstrated remarkable improvements over the [111–114], YOLOv5n, YOLOv5s, YOLOX tiny, YOLOv6n, YOLOv7 tiny, YOLOv8n, GELAN, YOLOv9, and YOLOv10n methods in mAP@0.5, surpassing other

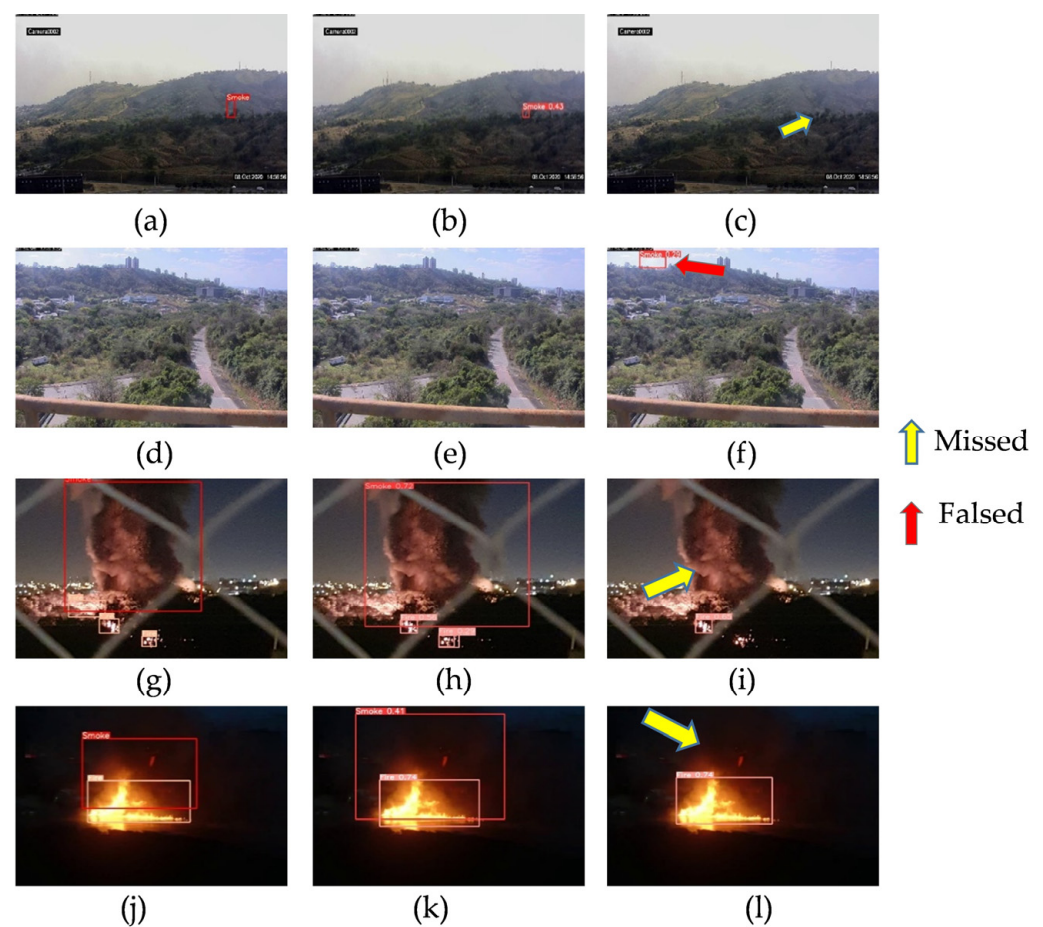
methods by 2.7%, 3.1%, 2.7%, 1.3%, 3.0%, 1.8%, 2.7%, 1.2%, 3.7%, 1.8%, 1.0%, 2.0%, 1.5%, 1.4%, and 2.3%, respectively.

Despite the slight increase in inference time and floating-point operations observed compared with the baseline model, resulting in a minor decrease in FPS, the significant gains in accuracy, with mAP@0.5, precision, and recall increasing by 2.0%, 2.3%, and 2.7%, respectively, highlighting the value of the tradeoff.

These substantial advancements firmly establish ESFD-YOLOv8n as a state-of-the-art solution for smoke and fire detection in complex environments, maintaining real-time performance while significantly enhancing accuracy.

### 3.7. Qualitative Analysis between ESFD-YOLOv8n and the Original YOLOv8n Model

Figure 12 shows the prediction results from the ESFD-YOLOv8n and original YOLOv8n models.



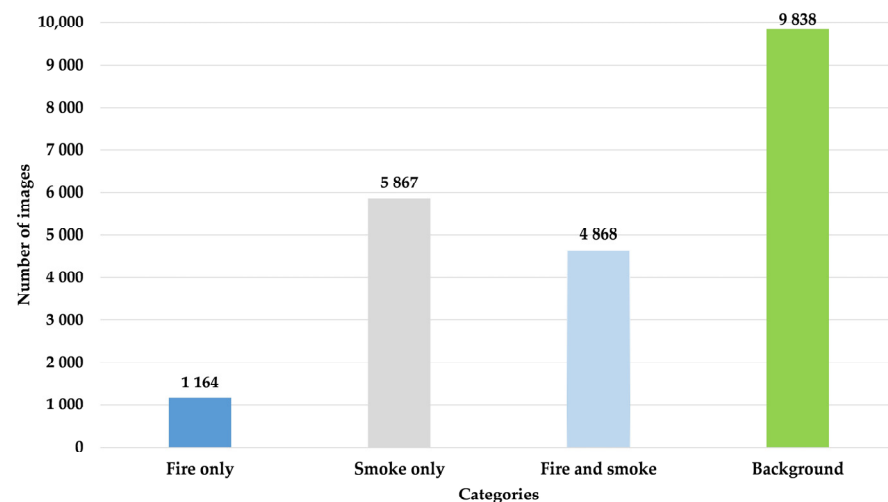
**Figure 12.** Detection results of the original and ESFD-YOLOv8n models. (a,d,g,j) Labeled images, (b,e,h,k) images predicted by ESFD-YOLOv8n, and (c,f,i,l) images predicted by YOLOv8n.

Figure 12 confirms the superior detection performance of ESFD-YOLOv8n in real-world fire detection applications compared with the original YOLOv8n model. In particular, Figure 12b shows how the enhanced model prioritizes relevant information and effectively preserves features crucial for detecting small smoke plumes, a critical aspect of early fire safety. Furthermore, the original YOLOv8n model showed instances of missed detections (false negatives, Figure 12c,i,l) and incorrect identifications (false positives, Figure 12f), issues that are reduced significantly in the ESFD-YOLOv8n model.

#### 4. Conclusions and Future Work

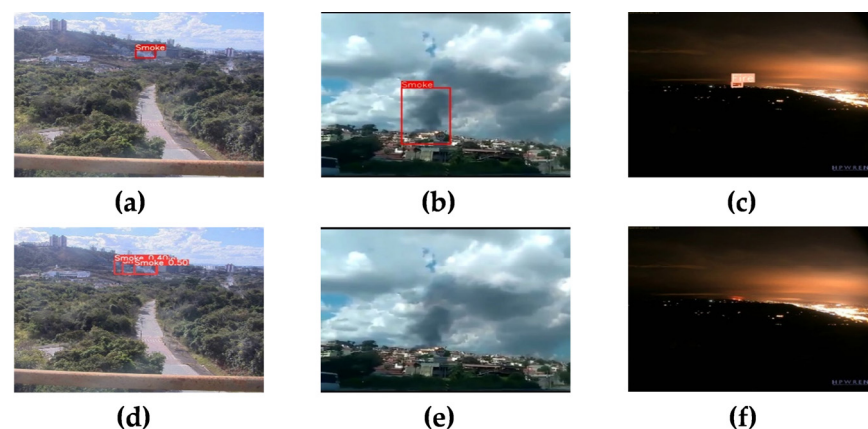
This paper introduced ESFD-YOLOv8n, a novel YOLOv8n-based model demonstrating significant improvements in smoke and fire detection accuracy and efficiency in early smoke and fire detection. Leveraging key technologies, such as ResBlock, WIoUv3 loss, C2fGELAN blocks, and transfer learning, ESFD-YOLOv8n outperformed the existing state-of-the-art methods on the D-Fire dataset, achieving an impressive mAP50 of 79.4% with a real-time inference time of 1.0 ms.

However, a thorough error analysis revealed several challenges. The dataset suffers from an imbalanced number of classes and a high number of background images (Figure 13). This imbalance leads to a relatively high difference in accuracy among classes, an increased false positive rate, and reduced generalization.



**Figure 13.** D-Fire dataset structure. Number of images in each category.

Additionally, model performance varies in different scenarios, including redundant boxes at far distances, lack of detection with similar backgrounds, and incorrect detection of similar objects, as shown in Figure 14.



**Figure 14.** Incorrect detections by the model. (a–c) Labeled images, (d) redundant boxes at far distances, (e) lack of detection with a similar background, and (f) incorrect detection of similar objects.

To address the issue of redundant boxes at far distances, it is recommended to implement NMS and adjust anchor box scales. For the lack of detection with similar backgrounds, using a diverse training dataset, applying data augmentation techniques such as random cropping and color jittering, exploring alternative loss functions, and addressing class imbalance are suggested. To improve detection accuracy for similar objects, increasing object class granularity and incorporating contextual information are recommended.

Overall, recommendations for enhancing the model include fine-tuning hyperparameters, regularly evaluating the model on a validation set, and considering ensemble methods. In our future work, we will address these limitations to further refine the model. By overcoming these challenges, we aim to develop a more reliable and versatile smoke and fire detection model for practical applications by enhancing accuracy and reducing false alarms. This advancement is expected to improve safety measures and early warning systems, ultimately mitigating fire hazards and protecting communities. The promising results of ESFD-YOLOv8n lay the groundwork for further research in this critical field, paving the way for more sophisticated fire detection systems across various applications.

**Author Contributions:** Conceptualization, D.M.; methodology, D.M. and P.L.M.T.; software, D.M.; validation, D.M., P.L.M.T., J.-H.K., and S.-C.K.; formal analysis, D.M.; investigation, D.M.; resources, D.M.; data curation, D.M.; writing—original draft preparation, D.M.; writing—review and editing, D.M., P.L.M.T., J.-H.K., and S.-C.K.; visualization, D.M. and P.L.M.T.; supervision, P.L.M.T., J.-H.K., and S.-C.K.; funding acquisition, S.-C.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study in this paper did not involve humans or animals.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Acknowledgments:** This research was supported by BK21PLUS, Creative Human Resource Development Program for IT Convergence. This research was supported by a 2-year Research Grant from Pusan National University, and by 2024 Specialization Project of Pusan National University.

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

## References

1. Fire Loss in the United States during 2022. Available online: <https://www.nfpa.org/education-and-research/research/nfpa-research/fire-statistical-reports/fire-loss-in-the-united-states> (accessed on 12 April 2024).
2. Burns. Available online: <https://www.who.int/news-room/fact-sheets/detail/burns> (accessed on 12 April 2024).
3. Khan, F.; Xu, Z.; Sun, J.; Khan, F.M.; Ahmed, A.; Zhao, Y. Recent Advances in Sensors for Fire Detection. *Sensors* **2022**, *22*, 3310. [CrossRef]
4. Lv, L.Y.; Cao, C.F.; Qu, Y.X.; Zhang, G.D.; Zhao, L.; Cao, K.; Song, P.; Tang, L.C. Smart fire-warning materials and sensors: Design principle, performances, and applications. *Mater. Sci. Eng. R Rep.* **2022**, *150*, 100690. [CrossRef]
5. Bao, Y.; Huang, Y.; Hoehler, M.S.; Chen, G. Review of Fiber Optic Sensors for Structural Fire Engineering. *Sensors* **2019**, *19*, 877. [CrossRef] [PubMed]
6. Gaur, A.; Singh, A.; Kumar, A.; Kulkarni, K.S.; Lala, S.; Kapoor, K.; Srivastava, V.; Kumar, A.; Mukhopadhyay, S.C. Fire Sensing Technologies: A Review. *IEEE Sens. J.* **2019**, *19*, 3191–3202. [CrossRef]
7. Gaur, A.; Singh, A.; Kumar, A.; Kumar, A.; Kapoor, K. Video Flame and Smoke Based Fire Detection Algorithms: A Literature Review. *Fire Technol.* **2020**, *56*, 1943–1980. [CrossRef]
8. Lee, Y.; Shim, J. False Positive Decrement Research for Fire and Smoke Detection in Surveillance Camera using Spatial and Temporal Features Based on Deep Learning. *Electronics* **2019**, *8*, 1167. [CrossRef]
9. Gagliardi, A.; Saponara, S. AdViSED: Advanced Video Smoke Detection for Real-Time Measurements in Antifire Indoor and Outdoor Systems. *Energies* **2020**, *13*, 2098. [CrossRef]
10. Bu, F.; Gharajeh, M.S. Intelligent and vision-based fire detection systems: A survey. *Image Vis. Comput.* **2019**, *91*, 103803. [CrossRef]
11. Nguyen, A.Q.; Nguyen, H.T.; Tran, V.C.; Pham, H.X.; Pestana, J. A Visual Real-time Fire Detection using Single Shot MultiBox Detector for UAV-based Fire Surveillance. In Proceedings of the 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE), Phu Quoc Island, Vietnam, 13–15 January 2021. [CrossRef]
12. Hosseini, A.; Hashemzadeh, M.; Farajzadeh, N. UFS-Net: A unified flame and smoke detection method for early detection of fire in video surveillance applications using CNNs. *J. Comput. Sci.* **2022**, *61*, 101638. [CrossRef]
13. Fonollosa, J.; Solórzano, A.; Marco, S. Chemical Sensor Systems and Associated Algorithms for Fire Detection: A Review. *Sensors* **2018**, *18*, 553. [CrossRef]
14. Benzekri, W.; El, A.; Moussaoui, O.; Berrajaa, M. Early Forest Fire Detection System using Wireless Sensor Network and Deep Learning. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 5. [CrossRef]



15. Ding, X.; Gao, J. A new intelligent fire color space approach for forest fire detection. *J. Intell. Fuzzy Syst.* **2022**, *42*, 5265–5281. [\[CrossRef\]](#)
16. Khalil, A.; Rahman, S.U.; Alam, F.; Ahmad, I.; Khalil, I. Fire Detection Using Multi Color Space and Background Modeling. *Fire Technol.* **2021**, *57*, 1221–1239. [\[CrossRef\]](#)
17. Dzigal, D.; Akagic, A.; Buza, E.; Brdjanin, A.; Dardagan, N. Forest Fire Detection based on Color Spaces Combination. In Proceedings of the 2019 11th International Conference on Electrical and Electronics Engineering (ELECO), Bursa, Turkey, 28–30 November 2019. [\[CrossRef\]](#)
18. Alamgir, N.; Nguyen, K.; Chandran, V.; Boles, W. Combining multi-channel color space with local binary co-occurrence feature descriptors for accurate smoke detection from surveillance videos. *Fire Saf. J.* **2018**, *102*, 1–10. [\[CrossRef\]](#)
19. 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. [\[CrossRef\]](#)
20. Jamali, M.; Karimi, N.; Samavi, S. Saliency Based Fire Detection Using Texture and Color Features. In Proceedings of the 2020 28th Iranian Conference on Electrical Engineering (ICEE), Tabriz, Iran, 4–6 August 2020. [\[CrossRef\]](#)
21. Luo, Y.; Zhao, L.; Liu, P.; Huang, D. Fire smoke detection algorithm based on motion characteristic and convolutional neural networks. *Multimed. Tools Appl.* **2018**, *77*, 15075–15092. [\[CrossRef\]](#)
22. Wu, X.; Lu, X.; Leung, H.A. Video Based Fire Smoke Detection Using Robust AdaBoost. *Sensors* **2018**, *18*, 3780. [\[CrossRef\]](#)
23. Islam, M.R.; Amiruzzaman, M.; Nasim, S.; Shin, J. Smoke Object Segmentation and the Dynamic Growth Feature Model for Video-Based Smoke Detection Systems. *Symmetry* **2020**, *12*, 1075. [\[CrossRef\]](#)
24. Geetha, S.; Abhishek, C.S.; Akshayanat, C.S. Machine Vision Based Fire Detection Techniques: A Survey. *Fire Technol.* **2021**, *57*, 591–623. [\[CrossRef\]](#)
25. Chaturvedi, S.; Khanna, P.; Ojha, A. A survey on vision-based outdoor smoke detection techniques for environmental safety. *ISPRS J. Photogramm. Remote Sens.* **2022**, *185*, 158–187. [\[CrossRef\]](#)
26. Boroujeni, S.P.H.; Razi, A.; Khoshdel, S.; Afghah, F.; Coen, J.L.; O'Neill, L.; Fule, P.; Watts, A.; Kokolakis, N.M.T.; Vamvoudakis, K.G. A comprehensive survey of research towards AI-enabled unmanned aerial systems in pre-, active-, and post-wildfire management. *Inf. Fusion.* **2024**, *108*, 102369. [\[CrossRef\]](#)
27. Saydirasulovich, S.N.; Mukhiddinov, M.; Djuraev, O.; Abdusalomov, A.; Young-Im, C. An Improved Wildfire Smoke Detection Based on YOLOv8 and UAV Images. *Sensors* **2023**, *23*, 8374. [\[CrossRef\]](#)
28. Jadon, A.; Omama, M.; Varshney, A.; Ansari, M.S.; Sharma, R. FireNet: A Specialized Lightweight Fire & Smoke Detection Model for Real-Time IoT Applications. *arXiv* **2019**, arXiv:1905.11922. [\[CrossRef\]](#)
29. Luo, W. Research on fire detection based on YOLOv5. In Proceedings of the 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Xi'an, China, 15–17 July 2022. [\[CrossRef\]](#)
30. Wang, Z.; Wu, L.; Li, T.; Shi, P.A. Smoke Detection Model Based on Improved YOLOv5. *Mathematics* **2022**, *10*, 1190. [\[CrossRef\]](#)
31. Li, J.; Xu, R.; Liu, Y. An Improved Forest Fire and Smoke Detection Model Based on YOLOv5. *Forests* **2023**, *14*, 833. [\[CrossRef\]](#)
32. Mukhiddinov, M.; Abdusalomov, A.B.; Cho, J. A Wildfire Smoke Detection System Using Unmanned Aerial Vehicle Images Based on the Optimized YOLOv5. *Sensors* **2022**, *22*, 9384. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Chen, X.; Xue, Y.; Zhu, Y.; Ma, R. A novel smoke detection algorithm based on improved mixed Gaussian and YOLOv5 for textile workshop environments. *IET Image Process.* **2023**, *17*, 1991–2004. [\[CrossRef\]](#)
34. Chiu, Y.C.; Tsai, C.Y.; Ruan, M.D.; Shen, G.Y.; Lee, T.T. Mobilenet-SSDv2: An Improved Object Detection Model for Embedded Systems. In Proceedings of the 2020 International Conference on System Science and Engineering (ICSSE), Takamatsu, Kagawa, Japan, 31 August–3 September 2020. [\[CrossRef\]](#)
35. Shi, Z.; Sun, R.; Huo, M. Smoke Video Detection Algorithm Based On 3D Convolutional Neural Network. In Proceedings of the 2022 34th Chinese Control and Decision Conference (CCDC), Hefei, China, 15–17 August 2022. [\[CrossRef\]](#)
36. Yu, F.; Yang, Q.; Zhang, G.; Jin, X.; Guo, D.; Wang, P.; Yao, G. An intelligent wildfire identification method based on weighted boxes fusion and convolutional block attention module. *Int. J. Parallel Emergent Distrib. Syst.* **2024**, 1–12. [\[CrossRef\]](#)
37. Dalal, S.; Lilhore, U.K.; Radulescu, M.; Simaiya, S.; Jaglan, V.; Sharma, A. A hybrid LBP-CNN with YOLO-v5-based fire and smoke detection model in various environmental conditions for environmental sustainability in smart city. *Environ. Sci. Pollut. Res.* **2024**. [\[CrossRef\]](#)
38. Xiao, Z.; Wan, F.; Lei, G.; Xiong, Y.; Xu, L.; Ye, Z.; Liu, W.; Zhou, W.; Xu, C. FL-YOLOv7: A Lightweight Small Object Detection Algorithm in Forest Fire Detection. *Forests* **2023**, *14*, 1812. [\[CrossRef\]](#)
39. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. GhostNet: More Features from Cheap Operations. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2020, Seattle, WA, USA, 13–19 June 2020. [\[CrossRef\]](#)
40. de Venâncio, P.V.A.B.; Campos, R.J.; Rezende, T.M.; Lisboa, A.C.; Barbosa, A.V. A hybrid method for fire detection based on spatial and temporal patterns. *Neural Comput. Appl.* **2023**, *35*, 9349–9361. [\[CrossRef\]](#)
41. Talaat, F.M.; ZainEldin, H. An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Comput. Appl.* **2023**, *35*, 20939–20954. [\[CrossRef\]](#)
42. de Venâncio, P.V.A.B.; Lisboa, A.C.; Barbosa, A.V. An automatic fire detection system based on deep convolutional neural networks for low-power, resource-constrained devices. *Neural Comput. Appl.* **2022**, *34*, 15349–15368. [\[CrossRef\]](#)
43. Khan, M.A.; Park, H. FireXplainNet: Optimizing Convolution Block Architecture for Enhanced Wildfire Detection and Interpretability. *Electronics* **2024**, *13*, 1881. [\[CrossRef\]](#)



44. Qiao, L.; Li, S.; Zhang, Y.; Yan, J. Early Wildfire Detection and Distance Estimation Using Aerial Visible-Infrared Images. *IEEE Trans. Ind. Electron.* **2024**, 1–11. [\[CrossRef\]](#)
45. Wang, G.; Li, H.; Li, P.; Lang, X.; Feng, Y.; Ding, Z.; Xie, S. M<sup>4</sup> SFWD: A Multi-Faceted synthetic dataset for remote sensing forest wildfires detection. *Expert. Syst. Appl.* **2024**, *248*, 123489. [\[CrossRef\]](#)
46. Jin, S.; Wang, T.; Huang, H.; Zheng, X.; Li, T.; Guo, Z. A self-adaptive wildfire detection algorithm by fusing physical and deep learning schemes. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *127*, 103671. [\[CrossRef\]](#)
47. Jonnalagadda, A.V.; Hashim, H.A. SegNet: A segmented deep learning based Convolutional Neural Network approach for drones wildfire detection. *Remote Sens. Appl.* **2024**, *34*, 101181. [\[CrossRef\]](#)
48. Liang, T.; Zeng, G. FSH-DETR: An Efficient End-to-End Fire Smoke and Human Detection Based on a Deformable DETection TRansformer (DETR). *Sensors* **2024**, *24*, 4077. [\[CrossRef\]](#)
49. Li, R.; Hu, Y.; Li, L.; Guan, R.; Yang, R.; Zhan, J.; Cai, W.; Wang, Y.; Xu, H.; Li, L. SMWE-GFPNNet: A high-precision and robust method for forest fire smoke detection. *Knowl. Based Syst.* **2024**, *289*, 111528. [\[CrossRef\]](#)
50. Zheng, H.; Wang, G.; Xiao, D.; Liu, H.; Hu, X. FTA-DETR: An efficient and precise fire detection framework based on an end-to-end architecture applicable to embedded platforms. *Expert. Syst. Appl.* **2024**, *248*, 123394. [\[CrossRef\]](#)
51. Yang, L.; Feng, Y.; Wang, Y.; Wang, J. Refined fire detection and band selection method in hyperspectral remote sensing imagery based on sparse-VIT. *Infrared Phys. Technol.* **2024**, *137*, 105104. [\[CrossRef\]](#)
52. Tong, H.; Yuan, J.; Zhang, J.; Wang, H.; Li, T. Real-Time Wildfire Monitoring Using Low-Altitude Remote Sensing Imagery. *Remote Sens.* **2024**, *16*, 2827. [\[CrossRef\]](#)
53. Thangavel, K.; Spiller, D.; Sabatini, R.; Amici, S.; Sasidharan, S.T.; Fayek, H.; Marzocca, P. Autonomous Satellite Wildfire Detection Using Hyperspectral Imagery and Neural Networks: A Case Study on Australian Wildfire. *Remote Sens.* **2023**, *15*, 720. [\[CrossRef\]](#)
54. Yun, B.; Zheng, Y.; Lin, Z.; Li, T. FFYOLO: A Lightweight Forest Fire Detection Model Based on YOLOv8. *Fire* **2024**, *7*, 93. [\[CrossRef\]](#)
55. Shaddy, B.; Ray, D.; Farguell, A.; Calaza, V.; Mandel, J.; Haley, J.; Hilburn, K.; Mallia, D.V.; Kochanski, A.; Oberai, A. Generative Algorithms for Fusion of Physics-Based Wildfire Spread Models with Satellite Data for Initializing Wildfire Forecasts. *Artif. Intell. Earth Syst.* **2024**, *3*, e230087. [\[CrossRef\]](#)
56. Jayagopal, P.; Purushothaman Janaki, K.; Mohan, P.; Kondapaneni, U.B.; Periyasamy, J.; Mathivanan, S.K.; Dalu, G.T. A modified generative adversarial networks with Yolov5 for automated forest health diagnosis from aerial imagery and Tabu search algorithm. *Sci. Rep.* **2024**, *14*, 4814. [\[CrossRef\]](#)
57. Boroujeni, S.P.H.; Razi, A. IC-GAN: An Improved Conditional Generative Adversarial Network for RGB-to-IR image translation with applications to forest fire monitoring. *Expert. Syst. Appl.* **2024**, *238*, 121962. [\[CrossRef\]](#)
58. Shawly, T.; Alsheikhy, A.A. Fire Identification Based on Novel Dense Generative Adversarial Networks. *Artif. Intell. Rev.* **2024**, *57*, 207. [\[CrossRef\]](#)
59. Chan, Y.W.; Liu, J.C.; Kristiani, E.; Lien, K.Y.; Yang, C.T. Flame and smoke detection using Kafka on edge devices. *Internet Things* **2024**, *27*, 101309. [\[CrossRef\]](#)
60. Yang, M.; Qian, S.; Wu, X. Real-time fire and smoke detection with transfer learning based on cloud-edge collaborative architecture. *IET Image Process.* **2024**. [\[CrossRef\]](#)
61. Lehnert, A.; Gawantka, F.; During, J.; Just, F.; Reichenbach, M. XplAIable: Explainable AI Smoke Detection at the Edge. *Big Data Cogn. Comput.* **2024**, *8*, 50. [\[CrossRef\]](#)
62. Peruzzi, G.; Pozzebon, A.; Van Der Meer, M. Fight Fire with Fire: Detecting Forest Fires with Embedded Machine Learning Models Dealing with Audio and Images on Low Power IoT Devices. *Sensors* **2023**, *23*, 783. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Ultralytics YOLOv8. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 17 June 2024).
64. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [\[CrossRef\]](#)
65. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [\[CrossRef\]](#)
66. Wang, C.Y.; Yeh, I.H.; Liao, H.Y.M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv* **2024**, arXiv:2402.13616. [\[CrossRef\]](#)
67. 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. [\[CrossRef\]](#)
68. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IOU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051. [\[CrossRef\]](#)
69. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [\[CrossRef\]](#)
70. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767. [\[CrossRef\]](#)
71. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [\[CrossRef\]](#)
72. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [\[CrossRef\]](#)
73. YOLOv5 by Ultralytics. Available online: <https://github.com/ultralytics/yolov5> (accessed on 17 June 2024).

74. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976. [\[CrossRef\]](#)
75. 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 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023. [\[CrossRef\]](#)
76. Wang, C.-Y.; Liao, H.-Y.M.; Yeh, I.-H. Designing Network Design Strategies Through Gradient Path Analysis. *arXiv* **2023**, arXiv:2211.04800. [\[CrossRef\]](#)
77. Wang, A.; Chen, H.; Liu, L.; Kai, C.; Zijia, L.; Jungong, H.; Guiguang, D. YOLOv10: Real-Time End-to-End Object Detection. *arXiv* **2024**, arXiv:2405.14458. [\[CrossRef\]](#)
78. Yin, Y.; Li, H.; Fu, W. Faster-YOLO: An accurate and faster object detection method. *Digit. Signal Process.* **2020**, *102*, 102756. [\[CrossRef\]](#)
79. Sahafi, A.; Koulaouzidis, A.; Lalinia, M. Polypoid Lesion Segmentation Using YOLO-V8 Network in Wireless Video Capsule Endoscopy Images. *Diagnostics* **2024**, *14*, 474. [\[CrossRef\]](#)
80. Yu, Z.; Huang, H.; Chen, W.; Su, Y.; Liu, Y.; Wang, X. YOLO-FaceV2: A scale and occlusion aware face detector. *Pattern Recognit.* **2024**, *155*, 110714. [\[CrossRef\]](#)
81. Sekharamanthy, P.K.; Melgani, F.; Malacarne, J.; Ricci, R.; de Almeida Silva, R.; Marcato Junior, J. A Seamless Deep Learning Approach for Apple Detection, Depth Estimation, and Tracking Using YOLO Models Enhanced by Multi-Head Attention Mechanism. *Computers* **2024**, *13*, 83. [\[CrossRef\]](#)
82. Liu, M.; Wang, X.; Zhou, A.; Fu, X.; Ma, Y.; Piao, C. UAV-YOLO: Small Object Detection on Unmanned Aerial Vehicle Perspective. *Sensors* **2020**, *20*, 2238. [\[CrossRef\]](#)
83. Pi, Y.; Nath, N.D.; Behzadan, A.H. Convolutional neural networks for object detection in aerial imagery for disaster response and recovery. *Adv. Eng. Inform.* **2020**, *43*, 101009. [\[CrossRef\]](#)
84. Yang, F.; Fan, H.; Chu, P.; Blasch, E.; Ling, H. Clustered Object Detection in Aerial Images. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019. [\[CrossRef\]](#)
85. Yang, H.; Wang, J.; Wang, J. Efficient Detection of Forest Fire Smoke in UAV Aerial Imagery Based on an Improved Yolov5 Model and Transfer Learning. *Remote Sens.* **2023**, *15*, 5527. [\[CrossRef\]](#)
86. Zhang, M.; Xu, S.; Song, W.; He, Q.; Wei, Q. Lightweight Underwater Object Detection Based on YOLO v4 and Multi-Scale Attentional Feature Fusion. *Remote Sens.* **2021**, *13*, 4706. [\[CrossRef\]](#)
87. Jalal, A.; Salman, A.; Mian, A.; Shortis, M.; Shafait, F. Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecol. Inf.* **2020**, *57*, 101088. [\[CrossRef\]](#)
88. Hu, X.; Liu, Y.; Zhao, Z.; Liu, J.; Yang, X.; Sun, C.; Chen, S.; Li, B.; Zhou, C. Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network. *Comput. Electron. Agric.* **2021**, *185*, 106135. [\[CrossRef\]](#)
89. Mbouembe, P.L.T.; Liu, G.; Sikati, J.; Kim, S.C.; Kim, J.H. An efficient tomato-detection method based on improved YOLOv4-tiny model in complex environment. *Front. Plant Sci.* **2023**, *14*, 1150958. [\[CrossRef\]](#) [\[PubMed\]](#)
90. Touko Mbouembe, P.L.; Liu, G.; Park, S.; Kim, J.H. Accurate and fast detection of tomatoes based on improved YOLOv5s in natural environments. *Front. Plant Sci.* **2024**, *14*, 1292766. [\[CrossRef\]](#)
91. Aly, G.H.; Marey, M.; El-Sayed, S.A.; Tolba, M.F. YOLO Based Breast Masses Detection and Classification in Full-Field Digital Mammograms. *Comput. Methods Programs Biomed.* **2021**, *200*, 105823. [\[CrossRef\]](#)
92. Ramachandran, S.; George, J.; Skaria, S.; Varun, V.V. Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans. In Proceedings of the Medical Imaging 2018: Computer-Aided Diagnosis, Houston, TX, USA, 10–15 February 2018. [\[CrossRef\]](#)
93. Yang, R.; Yu, Y. Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis. *Front. Oncol.* **2021**, *11*, 638182. [\[CrossRef\]](#)
94. Pedoeem, J.; Huang, R.; Chen, C. YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018. [\[CrossRef\]](#)
95. Cai, Y.; Li, H.; Yuan, G.; Niu, W.; Li, Y.; Tang, X.; Ren, B.; Wang, Y. YOLObile: Real-Time Object Detection on Mobile Devices via Compression-Compilation Co-Design. In Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI-21), Online, 2–9 February 2021. [\[CrossRef\]](#)
96. Fang, W.; Wang, L.; Ren, P. Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments. *IEEE Access* **2020**, *8*, 1935–1944. [\[CrossRef\]](#)
97. Li, Y.; Wang, H.; Dang, L.M.; Nguyen, T.N.; Han, D.; Lee, A.; Jang, I.; Moon, H. A Deep Learning-Based Hybrid Framework for Object Detection and Recognition in Autonomous Driving. *IEEE Access* **2020**, *8*, 194228–194239. [\[CrossRef\]](#)
98. Jing, L.; Yang, X.; Tian, Y. Video you only look once: Overall temporal convolutions for action recognition. *J. Vis. Commun. Image Represent.* **2018**, *52*, 58–65. [\[CrossRef\]](#)
99. Lee, Y.; Park, J. CenterMask: Real-Time Anchor-Free Instance Segmentation. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020. [\[CrossRef\]](#)
100. Wu, T.; Dong, Y. YOLO-SE: Improved YOLOv8 for Remote Sensing Object Detection and Recognition. *Appl. Sci.* **2023**, *13*, 12977. [\[CrossRef\]](#)

101. Wang, C.-Y.; Liao, H.-Y.M.; Yeh, I.-H.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W. CSPNet: A New Backbone that can Enhance Learning Capability of CNN. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020. [\[CrossRef\]](#)
102. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille, France, 6–11 July 2015. [\[CrossRef\]](#)
103. Elfving, S.; Uchibe, E.; Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Netw.* **2018**, *107*, 3–11. [\[CrossRef\]](#)
104. Wang, K.; Liew, J.H.; Zou, Y.; Zhou, D.; Feng, J. PANet: Few-Shot Image Semantic Segmentation with Prototype Alignment. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019. [\[CrossRef\]](#)
105. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. UnitBox: An Advanced Object Detection Network. In *Proceedings of the MM '16: Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016.* [\[CrossRef\]](#)
106. D-Fire: An Image Dataset for Fire and Smoke Detection. Available online: <https://github.com/gaiasd/DFireDataset> (accessed on 12 April 2024).
107. Padilla, R.; Netto, S.L.; da Silva, E.A.B. A Survey on Performance Metrics for Object-Detection Algorithms. In Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 1–3 July 2020. [\[CrossRef\]](#)
108. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [\[CrossRef\]](#)
109. Muhammad, M.B.; Yeasin, M. Eigen-CAM: Class Activation Map using Principal Components. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020. [\[CrossRef\]](#)
110. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430. [\[CrossRef\]](#)
111. Park, M.; Tran, D.Q.; Bak, J.; Park, S. Advanced wildfire detection using generative adversarial network-based augmented datasets and weakly supervised object localization. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *114*, 103052. [\[CrossRef\]](#)
112. Yang, J.; Zhu, W.; Sun, T.; Ren, X.; Liu, F. Lightweight forest smoke and fire detection algorithm based on improved YOLOv5. *PLoS ONE* **2023**, *18*, e0291359. [\[CrossRef\]](#) [\[PubMed\]](#)
113. Xu, F.; Zhang, X.; Deng, T.; Xu, W. An Image-Based Fire Monitoring Algorithm Resistant to Fire-like Objects. *Fire* **2023**, *7*, 3. [\[CrossRef\]](#)
114. Liu, Z.; Zhang, R.; Zhong, H.; Sun, Y. YOLOv8 for Fire and Smoke Recognition Algorithm Integrated with the Convolutional Block Attention Module. *Open J. Appl. Sci.* **2024**, *14*, 159–170. [\[CrossRef\]](#)
115. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021. [\[CrossRef\]](#)
116. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. DETRs Beat YOLOs on Real-time Object Detection. *arXiv* **2023**, arXiv:2304.08069. [\[CrossRef\]](#)

**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.