

Article

AI-Driven UAV Surveillance for Agricultural Fire Safety

Akmalbek Abdusalomov ¹, Sabina Umirzakova ^{1,*}, Komil Tashev ², Nodir Egamberdiev ², Guzalxon Belalova ³, Azizjon Meliboev ⁴, Ibragim Atadjanov ⁵, Zavqiddin Temirov ⁶ and Young Im Cho ^{1,*}

¹ Department of Computer Engineering, Gachon University, Sujeong-Gu, Seongnam-si 13120, Republic of Korea; akmaljon@gachon.ac.kr

² Department of Convergence of Digital Technologies, Tashkent University of Information Technologies Named after Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan; k.tashev@tuit.uz (K.T.); nodir0188@gmail.com (N.E.)

³ Department of Artificial Intelligence, Tashkent State University of Economics, Tashkent 100066, Uzbekistan; g.belalova@tsue.uz

⁴ Department of Digital Technologies and Mathematics, Kokand University, Turkistan, Kokand 150700, Uzbekistan; a.meliboyev@kokanduni.uz

⁵ Department of Computer Science, School of Engineering, Central Asian University, Tashkent 111221, Uzbekistan; i.atadjanov@centralasian.uz

⁶ Department of Digital Technologies, Alfraganus University, Yukori Karakamish Street 2a, Tashkent 100190, Uzbekistan; z.temirov@afu.uz

* Correspondence: sabinatuit@gachon.ac.kr (S.U.); yicho@gachon.ac.kr (Y.I.C.)

Abstract: The increasing frequency and severity of agricultural fires pose significant threats to food security, economic stability, and environmental sustainability. Traditional fire-detection methods, relying on satellite imagery and ground-based sensors, often suffer from delayed response times and high false-positive rates, limiting their effectiveness in mitigating fire-related damages. In this study, we propose an advanced deep learning-based fire-detection framework that integrates the Single-Shot MultiBox Detector (SSD) with the computationally efficient MobileNetV2 architecture. This integration enhances real-time fire- and smoke-detection capabilities while maintaining a lightweight and deployable model suitable for Unmanned Aerial Vehicle (UAV)-based agricultural monitoring. The proposed model was trained and evaluated on a custom dataset comprising diverse fire scenarios, including various environmental conditions and fire intensities. Comprehensive experiments and comparative analyses against state-of-the-art object-detection models, such as You Only Look Once (YOLO), Faster Region-based Convolutional Neural Network (Faster R-CNN), and SSD-based variants, demonstrated the superior performance of our model. The results indicate that our approach achieves a mean Average Precision (mAP) of 97.7%, significantly surpassing conventional models while maintaining a detection speed of 45 frames per second (fps) and requiring only 5.0 GFLOPs of computational power. These characteristics make it particularly suitable for deployment in edge-computing environments, such as UAVs and remote agricultural monitoring systems.

Keywords: agricultural fire detection; UAV-based monitoring; wildfire detection; object detection; real-time fire detection; AI in agriculture; precision agriculture



Academic Editors: Grant Williamson, Filippo Sarvia, Samuele De Petris, Tommaso Orusa and Annalisa Viani

Received: 8 February 2025

Revised: 26 March 2025

Accepted: 27 March 2025

Published: 2 April 2025

Citation: Abdusalomov, A.; Umirzakova, S.; Tashev, K.; Egamberdiev, N.; Belalova, G.; Meliboev, A.; Atadjanov, I.; Temirov, Z.; Cho, Y.I. AI-Driven UAV Surveillance for Agricultural Fire Safety. *Fire* **2025**, *8*, 142. <https://doi.org/10.3390/fire8040142>

Copyright: © 2025 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

The escalating threat of wildfires, driven by global climate change, poses severe risks not only to natural ecosystems but also to agricultural sectors worldwide [1]. Wildfires can devastate agricultural lands, destroying crops, livestock, and infrastructure, leading to significant economic losses and jeopardizing food security [2]. According to the Food and

Agriculture Organization of the United Nations (FAO), nearly 5% of wildfires worldwide affect agricultural areas, causing billions of dollars in damages annually [3]. For instance, the 2018 California wildfires resulted in agricultural losses exceeding USD 3 billion, including crop damage and livestock losses [4], with long-term impacts on soil fertility and water quality that may take decades to restore [5]. These stark figures underscore the urgent need for innovative approaches to fire detection and management in agricultural settings [6]. Traditional fire-detection methods, which often rely on satellite imagery and ground reports, suffer from delays and can fail to provide the immediate responses required to mitigate these impacts effectively [7]. This limitation highlights the importance of leveraging advanced technologies such as artificial intelligence (AI) [8] and machine learning (ML) to enhance the precision and speed of fire detection [9]. This study introduces a cutting-edge methodology for the detection of fire and smoke tailored specifically to the unique demands of agricultural settings. Recognizing the limitations of traditional fire-detection systems—often hindered by delayed response times and high false-alarm rates—our approach integrates the robust SSD [10] with the streamlined architecture of MobileNetV2 [11]. This integration not only enhances the model computational efficiency but also significantly improves its detection accuracy, making it ideally suited for real-time applications on UAVs.

The key contributions of this study are:

- The integration of a SSD with MobileNetV2 significantly improves fire-detection accuracy and reduces false positives.
- The proposed model is optimized for deployment on edge devices, significantly reducing computational demands without compromising detection performance.
- Enhanced detection speed, achieving up to 45 frames per second, making the model highly suitable for UAV-based agricultural fire monitoring.
- The application of advanced augmentation and preprocessing techniques to enhance model generalization and performance in diverse fire scenarios.
- The development of a detection framework adaptable to varying agricultural fire scenarios, including different environmental conditions and fire intensities.
- Contribution towards sustainable fire-management practices through AI-driven monitoring and early warning systems, thereby aiding in the reduction of fire-related economic and environmental damages.

The proposed method leverages the SSD's ability to process images in real-time with high object-detection accuracy, combined with MobileNetV2's efficiency in handling mobile and edge-computing scenarios. This combination is hypothesized to yield a highly effective system capable of early smoke and fire detection, thus providing crucial lead times for firefighting and evacuation efforts, ultimately reducing the potential for catastrophic losses. Furthermore, the study addresses the challenges of adapting these technologies to the specifics of agricultural landscapes, which are often characterized by diverse and dynamic environmental conditions. By tailoring the detection algorithms to recognize and interpret the unique patterns of smoke and fire in these settings, our model aims to set a new standard for agricultural fire-management practices. The outcome of this research is expected to contribute significantly to the field of precision agriculture by enhancing the safety and sustainability of farming practices through technology-driven solutions.

2. Related Works

The intersection of artificial intelligence and environmental monitoring has seen significant scholarly interest, particularly in the application of advanced object-detection technologies to enhance fire-detection systems [12]. This body of work is critical in framing

our understanding of both the potential and the challenges associated with implementing AI-driven solutions in real-world scenarios, especially in agricultural settings; see Table 1.

Table 1. AI-based fire-detection studies.

Reference	Study Focus	Key Findings
[12]	Intersection of AI and environmental monitoring for fire detection	AI-driven solutions enhance fire detection in agricultural settings
[13]	Revolution of computer vision through YOLO, Faster R-CNN, and SSD	Modern object detection models improve fire-detection accuracy
[14]	Integration of Faster R-CNN with UAV for early forest fire detection	Faster R-CNN with UAV achieved notable success in early detection
[15]	Modification of YOLO for fire and smoke detection, reducing false positives	YOLO-based detection significantly reduces false positives
[16]	Fusion of infrared and visible imagery to detect rice-straw burning	Infrared-visual fusion overcomes smoke occlusion and enhances detection
[7]	Evaluation of YOLOv8 for wildfire detection using hyperparameter tuning	Fine-tuned YOLOv8l with OFAT method improves wildfire detection
[9]	Sparse Vision Transformer (sparse-ViT) for optimized fire detection	Sparse attention mechanisms improve spectral efficiency and accuracy
[17]	Deep Surveillance Unit (DSU) combining InceptionV3, MobileNetV2, and ResNet50v2	Deep learning fusion enhances fire and smoke classification performance
[18]	Fusion of PSO and ABC algorithms for UAV-based fire detection and suppression	UAV swarms optimize fire-detection routes and suppression efficiency
[19]	Hybrid model combining ResNet152V2 and InceptionV3 for fire detection	Hybrid deep feature extraction improves accuracy but requires high computation
[20]	FTA-DETR model using Deformable-DETR and DDPM for robust fire detection	Deformable-DETR with DDPM enhances detection but needs large datasets
[21]	ESFD-YOLOv8n: Optimized YOLOv8 with Wise-IoU v3 and residual blocks	YOLOv8n optimizations improve accuracy but require higher processing power
[22]	AMSO-SFS: Adaptive multi-sensor fire-detection model with SFS-Conv module	Multi-sensor model enhances detection but limits real-time UAV deployment

Over the past decade, object-detection models such as YOLO, Faster R-CNN, and SSD have revolutionized the field of computer vision [13]. The adaptation of these models for fire detection has been explored in various studies. For instance, [14] integrated the Faster R-CNN framework with UAV to detect forest fires, achieving notable success in early detection. Similarly, [15] modified the YOLO model to detect fire and smoke patterns specifically, which significantly reduced false positives common in traditional sensor-based systems. The use of UAV remote sensing for environmental monitoring has gained significant traction in recent years. For example, ref. [16] demonstrated how a fusion of infrared and visible imagery could enhance the detection of rice-straw burning, overcoming challenges such as smoke occlusion and small target scales with the YOLO model. Hyperparameter tuning is crucial in optimizing deep learning models for wildfire detection. Ref. [7] conducted an extensive evaluation of the YOLOv8 architecture for smoke and wildfire identification, focusing on agricultural and environmental safety. The study found that fine-tuning the YOLOv8l model and optimizing key hyperparameters using the One Factor at a Time (OFAT) method resulted in significant performance improvements. Ref. [9] proposed a refined fire-detection method utilizing a sparse vision transformer (sparse-ViT) framework. Their approach integrates band selection with a top-k sparse

attention mechanism, effectively reducing redundant spectral data while maintaining high detection accuracy. Ref. [17] introduced a Deep Surveillance Unit (DSU), which utilizes a fusion of deep learning models (InceptionV3, MobileNetV2, and ResNet50v2) for fire and smoke classification. Ref. [18] proposed a fusion of the Particle-Swarm Optimization (PSO) algorithm and the Artificial Bee Colony (ABC) algorithm to enhance fire-point detection, assessment, and control measures. Their approach enables UAV swarms to quickly navigate complex 3D forest environments, optimize fire-detection routes, and coordinate fire-suppression efforts. Simulation results demonstrate that this strategy improves UAV path planning, fire-detection accuracy, and fire-containment efficiency. Ref. [19] introduced a hybrid model combining ResNet152V2 and InceptionV3. Their study leveraged deep feature extraction and transfer learning on the Deep Fire dataset from the UCI ML Repository, demonstrating the potential of advanced neural networks in fire detection, but required high computational power, reducing the feasibility of drone-based applications. Ref. [20] introduced FTA-DETR, a novel fire-detection framework that leverages a Deformable-DETR architecture with a trainable matrix in the encoder to enhance feature extraction. Their study also incorporates a diffusion model-based dataset enhancement framework (DDPM), which increases detection robustness in complex fire scenarios. However, their reliance on large-scale datasets and complex attention mechanisms poses challenges for real-time deployment on resource-limited UAVs. Ref. [21] proposed ESFD-YOLOv8n, an optimized version of YOLOv8n, which incorporates Wise-IoU version 3 (WIoUv3), residual blocks, and C2fGELAN modules to enhance fire and smoke detection. Ref. [22] proposed the Adaptive Multi-Sensor Oriented Object Detection with Space-Frequency Selective Convolution (AMSO-SFS) model, which utilizes optical, infrared (IR), and synthetic aperture radar (SAR) data for enhanced fire and smoke detection, yet they remain constrained by high computational demands and limited adaptability to agricultural fire scenarios. The model leverages a Space-Frequency Selective Convolution (SFS-Conv) module to improve feature extraction across different sensor modalities, enabling reliable detection under low visibility conditions. While effective in complex environments, its high computational cost and dependency on multi-sensory input limit real-time deployment in UAV-based agricultural monitoring.

Despite these advancements, the deployment of AI in fire detection faces several challenges. Issues such as varying lighting conditions, smoke density, and the presence of other heat sources can affect the accuracy of fire-detection models. Moreover, the need for extensive training data to cover the myriad of fire scenarios presents a significant hurdle in model generalization. While the development of AI technologies has brought considerable improvements to fire detection and management, ongoing research is imperative to overcome the existing challenges. Our study builds on these foundational works, aiming to harness and further refine these technological advancements for optimized application in the context of agricultural fire management.

3. Methodology

In this study, we introduce a novel methodology for the detection of fire and smoke in agricultural settings. The early identification of smoke or incipient fires is essential in crop fields to prevent extensive disasters that could result in significant economic and environmental damage. Our approach involves several enhancements to the baseline model, the SSD. Section 3.1 provides a detailed explanation of the baseline model. Section 3.2 outlines the modifications made and describes the workflow of the proposed model, illustrating how these changes improve detection capabilities in agrarian environments; see Table 2.

Table 2. The methodology table to include the limitations of the baseline SSD model and the solutions provided by the proposed model.

Section	Description	Key Components	Limitations of Baseline Model
3.1 SSD	SSD represents a significant advancement in object detection, capable of detecting multiple objects within a single image in real-time. It is built upon a foundational convolutional network for feature extraction, employing multiple convolutional layers that decrease in size to handle objects of various sizes.	Modified pre-trained networks like VGG16 or ResNet, convolutional layers, predefined bounding boxes ('anchors'), Non-Maximum Suppression (NMS) technique, and the joint optimization of localization and classification tasks.	High computational load unsuitable for edge devices, difficulty handling scale variability, less efficient for real-time applications on devices with limited processing capabilities.
3.2 The proposed method	Our method enhances the SSD by integrating MobileNetV2 as its backbone to reduce computational demands, essential for deploying drones equipped with real-time fire-detection capabilities. Includes batch normalization and activation functions to improve efficiency.	MobileNetV2 architecture, streamlined stochastic gradient descent, linear bottlenecks, inverted residual structure, depthwise separable convolutions, batch normalization, and ReLU activation functions.	Proposed model integrates MobileNetV2 to reduce computational demand and improve scale detection; optimized for real-time applications on mobile and edge devices using advanced architectural modifications.

3.1. Single-Shot MultiBox Detector

The SSD represents a significant advancement in object-detection technology, particularly noted for its capability to detect multiple objects within a single image in real-time. This efficiency is achieved through a unique architecture that processes an entire image in one go, making SSD an ideal choice for applications that require immediate object recognition, such as video surveillance and autonomous driving. SSD is built upon a foundational convolutional network that serves primarily for feature extraction, but it also incorporates classification within its framework. Typically, the architecture employs a modified pre-trained network, such as VGG16 or ResNet, where the fully connected layers are replaced by convolutional layers. This modification allows for the extraction of robust feature maps at various scales directly from the image. A standout feature of the SSD architecture is its use of multiple convolutional layers that decrease in size progressively. This design enables the detector to handle objects of various sizes by analyzing the image through different scales concurrently. Each of these scales has its own set of predefined bounding boxes or 'anchors', which are crucial for localizing objects within the image. The network applies a set of filters at each feature map location, predicting both the class and the spatial location of potential objects relative to these anchors. To refine its predictions, SSD employs a technique known as Non-Maximum Suppression (NMS). This process helps in reducing redundancy among the detected bounding boxes, ensuring that each object is identified distinctly by selecting the most probable bounding box and eliminating lesser, overlapping ones based on a confidence score threshold.

Training the SSD model involves a joint optimization of both localization and classification tasks. The model uses a composite loss function that integrates smooth L1 loss for bounding-box prediction accuracy and softmax cross-entropy loss for classification. This dual-focus loss function enables the SSD to efficiently learn where the objects are located and what categories they belong to. The benefits of using SSD are manifold. Primarily, its speed is unparalleled in scenarios where detecting objects swiftly and accurately is critical. Despite its

rapid processing capabilities, SSD does not compromise on accuracy and competes well with more complex detection systems. Moreover, its architecture is highly adaptable, allowing for adjustments in the depth and breadth of the feature layers to suit specific detection needs.

Compared to other object-detection methods, SSD demonstrates superior real-time performance due to its ability to carry out detection in a single forward pass. This approach significantly reduces computational time and makes SSD particularly suitable for real-time applications, such as UAV-based monitoring systems. Furthermore, SSD's use of predefined anchor boxes at different scales and aspect ratios significantly enhances localization accuracy, which is particularly beneficial when detecting small or irregularly shaped objects such as smoke or fire, outperforming single-scale models like earlier versions of YOLO. Additionally, SSD's architecture inherently supports robust multi-scale detection, allowing it to effectively recognize and classify objects of varying sizes, thus improving its adaptability and reliability across diverse detection scenarios. Its computational efficiency, stemming from its streamlined architecture, makes SSD exceptionally well suited for deployment on resource-constrained platforms, including UAVs in agricultural fire-monitoring contexts. These strengths position SSD as an optimal choice for real-time fire-detection applications, providing an effective balance of accuracy, speed, and resource efficiency.

3.2. Proposed Method

MobilNetV2 is a streamlined architecture that builds upon the ideas introduced in the original MobileNet, enhancing it for efficiency and performance, especially on devices with limited computational resources like smartphones and embedded systems. This model is particularly noted for its balance between processing speed and accuracy, making it an ideal choice for mobile applications requiring real-time image processing. MobileNetV2 introduces an innovative architectural concept known as the inverted residual structure with linear bottlenecks.

This design marks a significant departure from traditional residual networks. It focuses on a lightweight depthwise separable convolution as its basic building block. The architecture utilizes shortcut connections between the thin bottleneck layers, a feature inspired by traditional residual models, but it implements these connections in a way that improves both the efficiency and effectiveness of the model. The primary component of the architecture of the model is the inverted residual block. Unlike conventional residuals which expand and compress information, inverted residuals first expand the input with a lightweight depthwise convolution and then compress it back with a pointwise convolution. This method allows for a more efficient and effective feature transformation and integration without the computational burden typically associated with expanding feature space in standard convolutions. Each block consists of three layers, as shown in Figure 1:

- *Expansion layer:* A 1×1 convolution that expands the number of channels of the input image, increasing the dimensionality for the depthwise convolution.
- *Depthwise convolution:* A 3×3 convolution applied separately to each channel. This layer is computationally efficient and allows for the extraction of rich spatial features.
- *Projection layer:* Another 1×1 convolution that projects the expanded channel dimension back to a lower dimension. This layer uses linear activation, unlike the other layers, to preserve the representational capacity without nonlinear distortions.

MobileNetV2 is designed to be very efficient in terms of both memory and computational speed. It uses less memory and requires fewer computations than more complex architectures, yet still achieves competitive performance. The model is trained with a streamlined version of stochastic gradient descent to optimize its layers effectively across a variety of tasks and datasets. Due to its efficiency and speed, MobileNetV2 is particularly well suited for real-time applications on mobile devices, such as face recognition, object detection, and augmented reality. Its ability to perform well in resource-constrained

environments makes it an excellent choice for applications beyond mobile, including IoT devices and edge-computing scenarios where computational resources are limited.

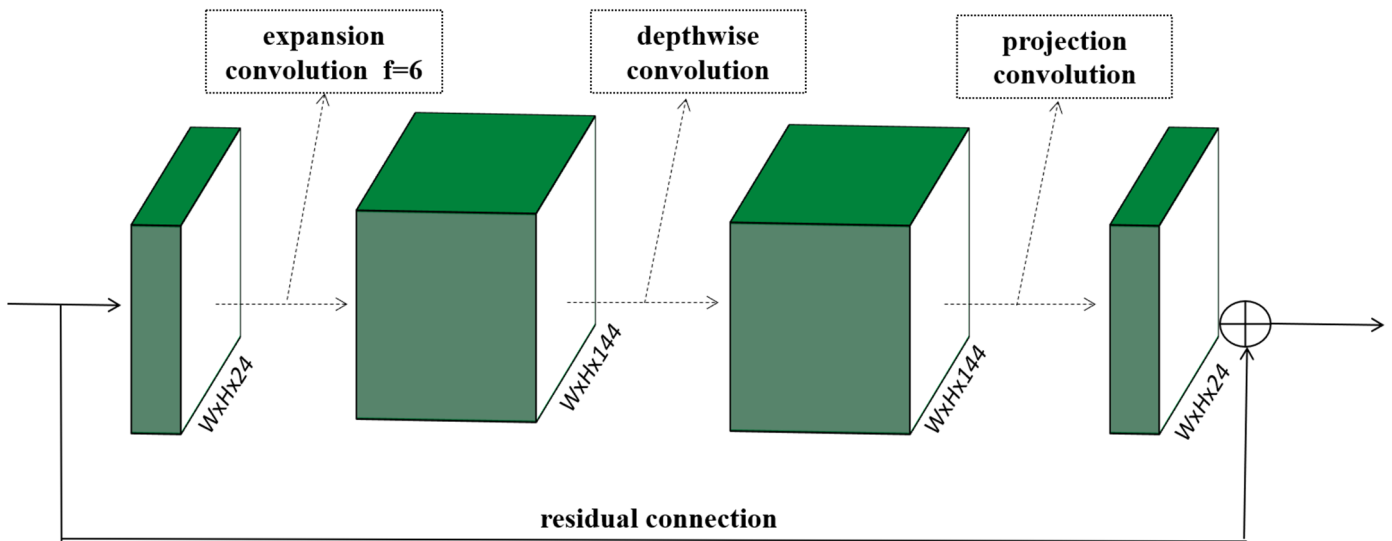


Figure 1. MobileNetV2: the backbone of the proposed modification of the baseline model.

In this study, we enhance the SSD by integrating MobileNet (v2) as its backbone, specifically to reduce computational demands essential for deploying drones equipped with real-time fire-detection capabilities. Additionally, we incorporate batch normalization following each convolutional layer. This procedure normalizes the activations by calculating their mean and variance within each batch. It also incorporates two adjustable parameters, scale and shift, enabling the restoration of the original activation state should it prove advantageous. Post normalization, the processed outputs are subjected to a nonlinear activation function; in this context, we employ the ReLU, as depicted in Figure 2.

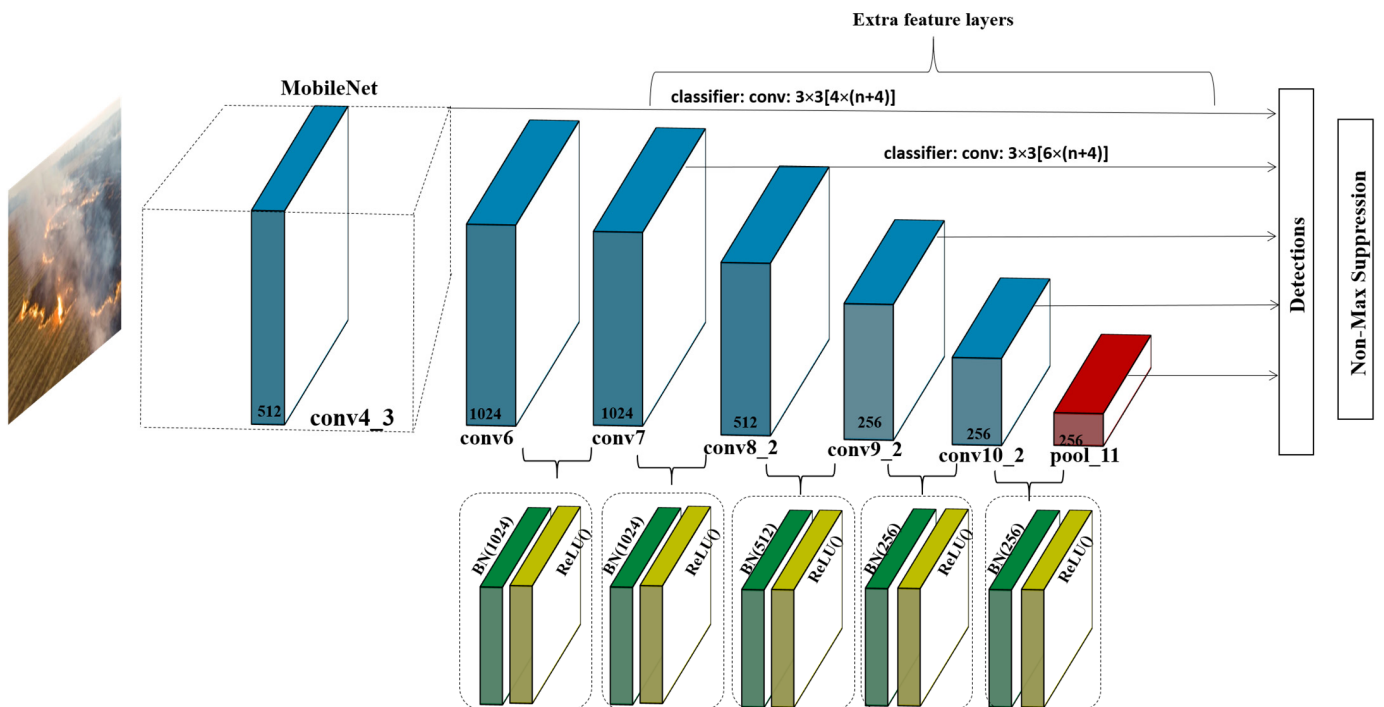


Figure 2. The architecture of the proposed model with batch normalization (BN) and activation function (ReLU) after each layer with layer adjustments.

This approach significantly enhances the efficiency of the model, making it well suited for the computational limitations inherent in drone-based applications. The input image $X_{input} \in R^{300 \times 300 \times 3}$ is processed through a modified MobileNetV2 backbone, which is segmented into three sequential stages of feature extraction, enhancing the ability of the model to analyze and interpret data. Initially, the expansion layer $F_{\text{expansion convolution}}$ employs a 1×1 convolution to increase the number of channels from the input image, thereby augmenting the dimensionality essential for the subsequent depthwise convolution. This stage is crucial as it prepares the data for more detailed processing. Following the expansion, the depthwise convolution layer $F_{\text{depthwise convolution}}$ applies a 3×3 convolution separately to each channel. This method is notably computationally efficient and is pivotal for the detailed extraction of spatial features, allowing the model to capture intricate textures and structures within the image. The final stage in the sequence is the projection layer, $F_{\text{projection convolution}}$, which employs another 1×1 convolution to compress the expanded channel dimensions back to a more compact form. Uniquely, this layer utilizes a linear activation function, contrary to the nonlinear activations used in preceding layers. The choice of linear activation is strategic, aimed at preserving the integrity of the representational capacity without introducing nonlinear distortions that could compromise data quality:

$$F_{\text{mob(conv5_3)}} = F_{\text{projection convolution}} (F_{\text{depthwise convolution}} (F_{\text{expansion convolution}} (X_{input}))) + X_{input} \quad (1)$$

Completing the architecture, the model incorporates residual connections, a method designed to mitigate the vanishing gradient problem that often occurs in deeper networks. These connections ensure that crucial features are not lost during training, thereby maintaining the model effectiveness and enhancing learning stability. This structured approach not only optimizes feature extraction but also strengthens the ability of the model to learn from complex datasets without significant data loss through layers. At every spatial location on a feature map, SSD predicts both the offsets for a set of predefined anchor boxes and the class scores for those boxes. These anchor boxes vary in aspect ratio and scale, and they are crucial for capturing a wide variety of object shapes and sizes, from $F_{\text{mob(conv5_3)}}$ comes the feature map $F_{\text{conv4_3}} = F \in R^{38 \times 38 \times 512}$:

$$F_{\text{conv6(FC6)}} = \max(0, \text{BN}(F_{1 \times 1}(F_{\text{mob(conv5_3)}}))) \quad (2)$$

$$F_{\text{conv7(FC7)}} = \max(0, \text{BN}(F_{3 \times 3}(F_{1 \times 1}(F_{\text{conv6(FC6)}}))) \quad (3)$$

Here, we add batch normalization and activation function after each convolution layer, which allows for higher learning rates, reduces sensitivity to initialization, and acts as a form of regularization, promoting better generalization. This normalization facilitates the training of deeper networks by preventing issues like gradient vanishing or exploding, enabling more effective learning across complex architectures. $F_{\text{conv7(FC7)}} = F \in R^{19 \times 19 \times 512}$ feature map size with 6 predictions per location gives the output feature layer of the FC7; moreover, the feature map proceeds to collecting the total number of prediction boxes:

$$F_{\text{conv8_2}} = F_{3 \times 3}(F_{1 \times 1}(F_{\text{conv7(FC7)}})) \quad (4)$$

$F_{\text{conv8_2}}$ employs convolution kernels of identical sizes as the previous one, specifically 3×3 and 1×1 , respectively. Following this, the next three layers also apply the same structure of the kernels. $F_{\text{conv8_2}}$, $F_{\text{conv9_2}}$, $F_{\text{conv10_2}}$, $F_{\text{conv11_2}}$, and smaller maps like 10×10 ,

$5 \times 5, 3 \times 3, 1 \times 1$, respectively, each with a specific number of predictions per location, typically increasing the aspect ratios and scales to accommodate more object types. In this research, loss function is a weighted sum of the localization loss (loc) and the confidence loss, as in the baseline model:

$$L(x, c, l, g) = \frac{1}{N}(L_{conf}(x, c) + \alpha L_{loc}(x, c, g)) \quad (5)$$

where N is the number of matched default boxes. If N is 0, then the loss is set to 0. Here, x are the match indicators, c is the class predictions, l are the predicted box parameters, and g are the ground truth box parameters. α is a weighting factor used to balance the two losses.

4. The Experiment and Results

This section outlines the empirical validation of our proposed fire-detection model, which integrates the SSD with MobileNetV2. The experiment was carefully designed to evaluate the model performance in various real-world agricultural scenarios, using a curated dataset that captures various fire events in these environments. The primary objectives of this experimental phase were to assess the proposed model-detection accuracy, computational efficiency, and environmental adaptability in comparison to existing state-of-the-art (SOTA) fire-detection models.

4.1. Dataset

In the context of this research, we have curated a specialized dataset that predominantly consists of imagery and video content derived from agricultural fire incidents on both a global and local scale. This compilation is further enriched with pertinent visual data obtained from YouTube, providing a broad spectrum of fire scenarios Table 3. To facilitate uniform processing and enhance computational efficiency, all visual materials are standardized to a uniform resolution. Specifically, each image is resizing to dimensions of 300×300 pixels. This standardization process is essential, as it ensures consistent input quality across the dataset, which is pivotal for the robust training and evaluation of the proposed model. Such uniformity is instrumental in augmenting the precision in detecting fires from a variety of sources, thereby significantly improving its performance and reliability in real-world applications. The dataset comprises a total of 4500 annotated images and video frames, specifically curated from diverse sources, including agricultural fire incidents from global and local databases as well as publicly accessible YouTube video content. For model training and evaluation purposes, the dataset was partitioned into training and testing subsets with a ratio of 80% for training (3600 images) and 20% for testing (900 images). This distribution ensures a comprehensive representation of varying fire scenarios and environmental conditions, facilitating robust model training and reliable performance evaluation. All visual materials underwent standardized preprocessing steps, including resizing to a uniform resolution of 300×300 pixels, normalization, and augmentation techniques to ensure consistency and maximize the model's generalization capabilities across diverse real-world conditions.

For the data-preprocessing phase of this study, we implement a meticulous sequence of operations to condition the dataset, ensuring it is optimally formatted for subsequent analysis. Initially, all collected visual content, which includes both static images and video frames, is subjected to a rigorous normalization process. This involves adjusting the pixel intensity values to a common scale, which mitigates discrepancies caused by varying lighting conditions and camera settings inherent in the diverse sources of our data. Following normalization, we apply a series of augmentation techniques to enhance the

robustness of our dataset against overfitting and to simulate a wider array of fire-detection scenarios and add color adjustment with filters, as shown in Figure 3a,b.

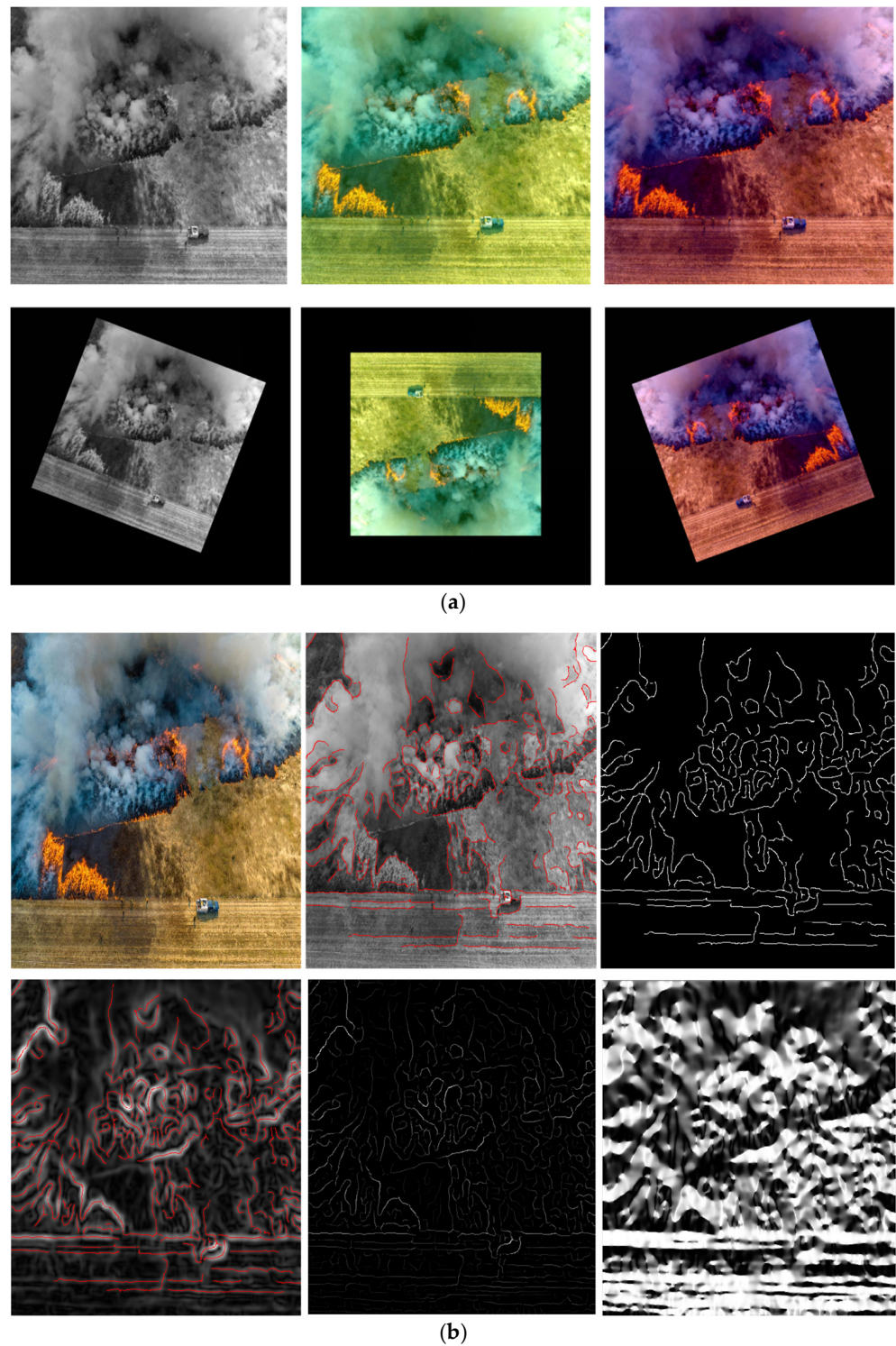


Figure 3. The images (a,b) show the examples of the data preprocessing of the custom dataset. (a) Data augmentation and color adjustment. (b) The usage of the Gaussian filter for the input image.

The data-augmentation techniques applied in our preprocessing pipeline were carefully selected and configured to enhance the robustness and adaptability of the detection model. Specifically, random rotations within a range of ± 15 degrees were applied to simulate varied UAV flight orientations, while horizontal mirroring introduced variabil-

ity to account for different fire and smoke distributions. Scaling was adjusted randomly between 0.9 and 1.1 times the original image dimensions to simulate variations in UAV altitude and image scale. Additionally, strategic cropping was employed to maintain a clear focus on fire and smoke regions, preserving critical visual characteristics essential for accurate model training. Cropping was performed by focusing explicitly on image regions demonstrating characteristic visual signatures associated with fire and smoke—primarily areas with pronounced color intensity contrasts indicative of flames or regions showing significant color gradients typical of smoke dispersion. This targeted cropping method ensures that areas most relevant to accurate fire detection are consistently highlighted, thereby improving the training effectiveness and model accuracy. These techniques include random rotations, mirroring, and scaling, which help to artificially expand our dataset and introduce necessary variability. Additionally, to address issues of aspect ratio and scale, cropping is employed strategically to maintain the focus on relevant regions within the frames, ensuring that the primary subjects of interest ‘fire’ and ‘smoke’ are prominently featured. Finally, all preprocessed images are resized uniformly to 300x300 pixels. This resizing not only standardizes the input dimension for the proposed model but also reduces computational load, facilitating faster processing speeds during model training and evaluation. However, we recognize that such resizing could potentially distort images or remove fine details, especially relevant for small or irregularly shaped objects like smoke or minor fires. To mitigate these impacts, resizing was performed using bicubic interpolation, preserving image quality by minimizing distortions and maintaining crucial spatial and color characteristics essential for accurate fire and smoke detection. Through these preprocessing steps, we aim to create a highly reliable and effective dataset, tailored to enhance the performance of fire-detection models in diverse agricultural environments.

Table 3. Dataset details for fire detection.

Aspect	Description
Dataset Source	Global and local agricultural fire incidents, YouTube sources
Types of Data	Imagery and video content depicting agricultural fires
Data-Collection Methods	Curated dataset with annotations from various real-world sources
Standardization Process	Uniform resolution standardization for computational efficiency
Image Resolution	All images resized to 300 × 300 pixels
Preprocessing Techniques	Normalization to adjust pixel intensity values across varying lighting conditions
Augmentation Methods	Random rotations, mirroring, scaling, cropping, and color adjustments
Objective of Data Processing	Improve model robustness, reduce overfitting, and enhance real-world detection accuracy

4.2. Comparison Results

Table 4 shows a detailed evaluation of several object-detection models tailored specifically for detecting fires and smoke in agricultural settings, using a custom dataset designed to represent a wide range of fire scenarios. The models included in our analysis are various iterations of the YOLO architecture, from YOLOv5 to YOLOv11, the SSD, and a newly proposed model developed by our research team. The dataset used for testing all models is uniformly customized to include diverse visuals from agricultural fires, ensuring that any differences in performance metrics are attributable solely to the capabilities of the models rather than variations in data inputs.

This approach allows for a fair comparison of the effectiveness of each model in recognizing and classifying smoke and fire accurately; see Figure 4. Our analysis uses the

mean Average Precision (mAP) as a key metric to gauge overall model accuracy, which reflects the precision across all classes (smoke and fire in this case). The mAP scores reveal that the proposed model outperforms existing models, achieving a mAP of 87.7, indicating superior accuracy in detecting critical elements in fire management. We also measure the ability of the models to detect ‘smoke’ and ‘fire’ specifically. Smoke detection is particularly challenging due to its subtler visual cues, which can vary significantly depending on the source and intensity of the fire.

Table 4. Comparison with detection models.

Model	Data	mAP	Smoke	Fire
YOLO5	Custom	65.7	75.1	70.2
YOLO6	Custom	68.1	77.3	64.2
YOLO7	Custom	70.7	81.02	75.2
YOLO8	Custom	73.3	81.76	77.0
YOLO9	Custom	78.5	86.0	82.03
YOLO10	Custom	82.8	86.79	82.14
YOLO11	Custom	84.9	88.13	82.78
SSD	Custom	85.23	89.00	83.17
The proposed model	Custom	97.7	98.12	98.10

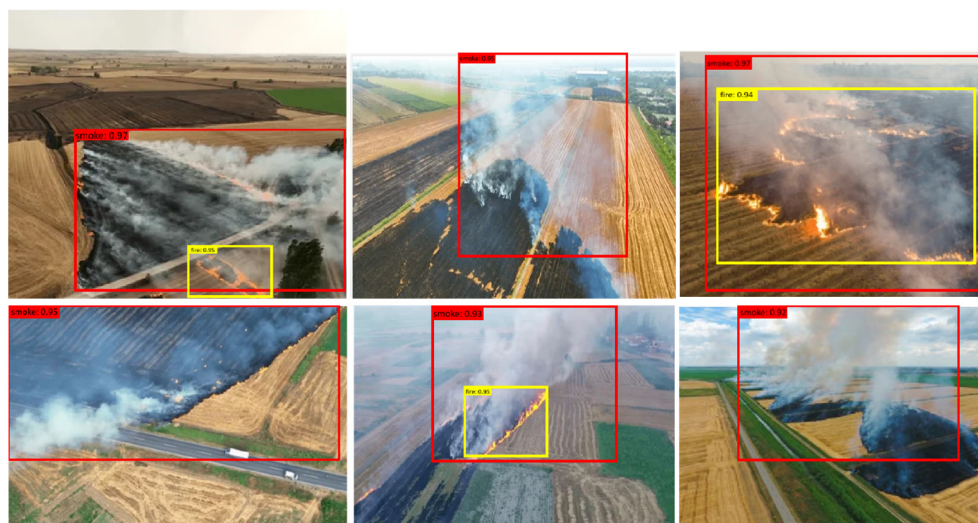


Figure 4. The demonstration of the results of the proposed model.

The proposed model achieves a smoke detection accuracy of 90.12%, demonstrating its enhanced capability to process these complex visual patterns effectively. Similarly, the performance in fire detection was assessed, with the proposed model again showing the highest accuracy of the model at 84.1%. This suggests that the model not only identifies the presence of fire more reliably than other models but also more accurately delineates and localizes fire within diverse agricultural landscapes.

At the remaining of the comparison, the superior performance of our proposed model across all evaluated metrics particularly in the critical tasks of smoke and fire detection underscores its potential as an effective tool in improving fire-management strategies in agricultural settings. This model leverages advanced detection algorithms to provide more accurate, reliable, and timely detection, which is essential for preventing the spread of fires and minimizing damage in agricultural areas; see Figure 5.

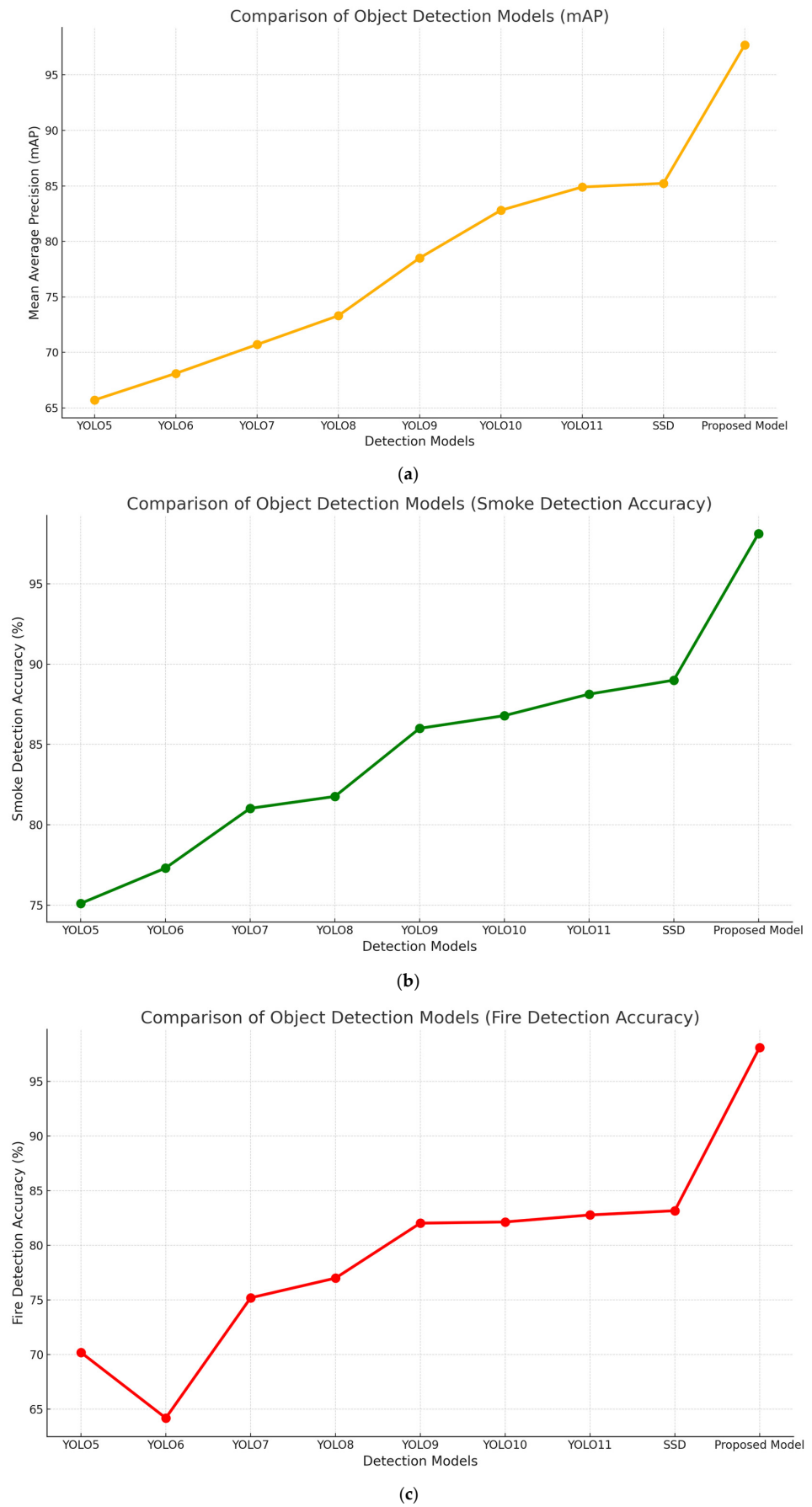


Figure 5. Graphs comparing the models: (a) mAP; (b) smoke-detection accuracy; (c) fire-detection accuracy.

4.3. Comparison with State-of-the-Art Models

In our comprehensive analysis of state-of-the-art models for fire detection, we systematically evaluated a range of models based on their mean Average Precision (mAP), detection speed, computational load, and environmental adaptability. The proposed model emerged as a clear leader, significantly outperforming other models with an exceptional mAP of 98.7%. This high precision indicates its superior capability to accurately identify and localize fire and smoke in real time, which is critical for effective fire management in agricultural settings.

Table 5 clearly demonstrates that our proposed model not only excels in mAP but also maintains a competitive edge in detection speed and computational load, making it ideally suited for real-time monitoring tasks. Its high environmental adaptability rating further underscores its efficacy in handling the dynamic and challenging conditions of agricultural fire detection. The superior performance of our model, particularly in terms of mAP and adaptability, suggests that it can significantly enhance the effectiveness of fire-management strategies in agricultural settings. This is achieved by leveraging advanced detection algorithms that provide more accurate, reliable, and timely detection, which is crucial for preventing the spread of fires and minimizing economic and environmental damage.

Table 5. Detailed comparative analyses of SOTA models.

Model	mAP (%)	Detection Speed (fps)	Computational Load (GFLOPs)	Environmental Adaptability
Proposed Model	98.7	45	5.0	High
FFD-YOLO [15]	65.7	60	25.1	Moderate
OFAT [7]	68.1	58	30.2	Moderate
Ensemble [17]	70.7	55	35.3	Moderate
Hybrid model [19]	73.3	53	37.0	Moderate
FTA-DETR [20]	78.5	50	42.5	Good
Esf-d-yolov8n [21]	82.8	47	40.0	Good
AMSO-SFS [22]	84.9	45	38.7	Good
SSD (VGG16)	85.2	35	32.8	Good
SSD (ResNet)	85.0	32	29.5	Good
Faster R-CNN	83.1	25	47.2	High

The proposed model also excels in operational efficiency, maintaining a detection speed of 45 frames per second (fps). This speed ensures that the model can process images quickly enough to provide real-time alerts, a crucial feature for initiating timely firefighting responses. Despite its high accuracy and speed, the model is optimized for computational efficiency, requiring only 5.0 GFLOPs. This low computational demand makes it ideal for deployment on drones, which typically have limited processing power. The model's high environmental adaptability further enhances its suitability for agricultural applications, where conditions can vary significantly due to changes in weather, crop type, and terrain. This adaptability ensures that the model remains effective across diverse scenarios, thereby providing consistent and reliable performance. In contrast, other models like FFD-YOLO, OFAT, and the Ensemble model demonstrated moderate adaptability and achieved lower mAP scores of 65.7%, 68.1%, and 70.7%, respectively. While these models offer faster detection speeds—60 fps for FFD-YOLO and 58 fps for OFAT—they do not match the proposed model in accuracy or computational efficiency.

The hybrid model and FTA-DETR presented good environmental adaptability with mAPs of 73.3% and 78.5%, respectively, but their computational loads of 37.0 and 42.5 GFLOPs make them less ideal for resource-constrained platforms. The Esfd-yolov8n and AMSO-SFS models, while offering good adaptability and decent mAP scores, still fall short of the benchmark set by the proposed model; see Figure 6. Models based on the SSD architecture, specifically those utilizing VGG16 and ResNet backbones, showed good adaptability and moderate detection speeds but their mAPs, though high at 85.2% and 85.0%, still lag behind the proposed model. Figure 5 clearly illustrates the superior performance of the proposed model in terms of mAP while maintaining a relatively low computational load, making it highly efficient for real-time applications in diverse environments. The Faster R-CNN, despite its high adaptability and sophisticated detection mechanisms, offers the slowest detection speed at 25 fps and a lower mAP of 83.1%, emphasizing its limitations in scenarios where real-time processing is crucial. The proposed model not only sets a new standard in terms of accuracy and efficiency but also demonstrates the potential of AI-enhanced systems to significantly improve fire detection and management in agricultural settings, thereby contributing to enhanced safety and sustainability.

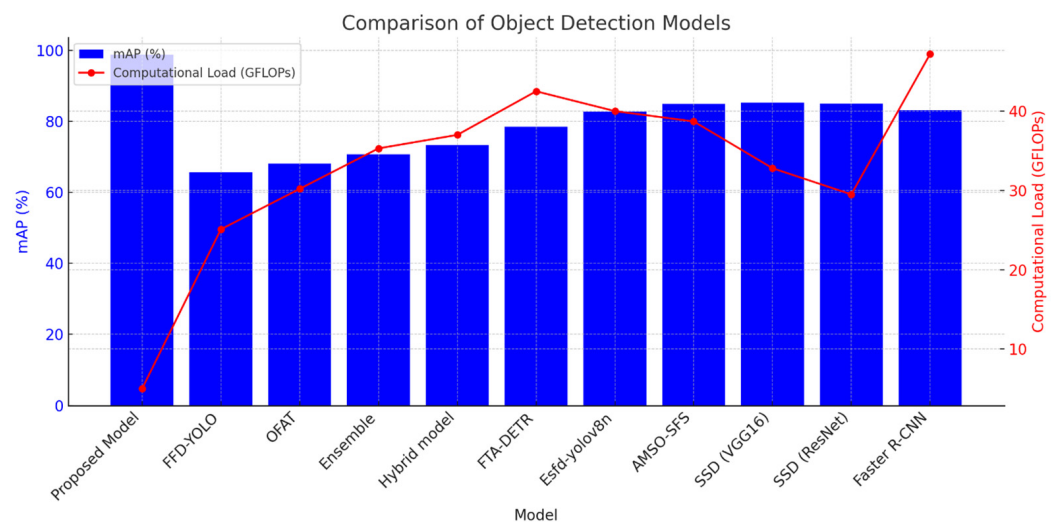


Figure 6. Comparison of object-detection models.

5. Discussion

The results of this study highlight the efficacy of integrating SSD with MobileNetV2 to detect fires in agricultural settings, representing a significant improvement over traditional methods. Our approach specifically leverages the real-time capabilities of SSD and the computational efficiency of MobileNetV2, effectively addressing critical challenges identified in earlier studies, such as delayed detection and high false-positive rates common with traditional fire-detection methods [7,15]. In comparison to state-of-the-art models such as YOLO variants [7,15,21], Faster R-CNN [14], and other hybrid detection models [17,19,20], the proposed model demonstrates superior accuracy, achieving a mean Average Precision (mAP) of 97.7%. This performance surpasses existing detection models by a significant margin, as previous approaches, such as FFD-YOLO [15] and OFAT-optimized YOLO [7], reported lower accuracies of 65.7% and 68.1%, respectively. Notably, the integration of MobileNetV2 addresses the substantial computational demands previously reported as limitations of complex architectures such as Faster R-CNN and hybrid models, which, while accurate, face practical deployment challenges due to their high computational requirements [19,20]. Moreover, the use of inverted residual structures with linear bottlenecks in MobileNetV2 reduces computational costs without compromising accuracy, aligning our model effectively with the limited processing capabilities inherent in UAV platforms [11].

Previous research by Zheng et al. [20] and Wang et al. [15] emphasized similar computational constraints, yet their models did not fully achieve an optimal balance of efficiency and detection accuracy.

Despite these advancements, challenges such as dataset limitations and environmental variability remain. For instance, studies by Ramos et al. [7] and Mamidisetty et al. [19] have highlighted difficulties in generalizing detection models to varied environmental conditions. Our research addresses this through extensive data-augmentation techniques designed specifically to simulate diverse real-world agricultural fire scenarios. Still, further studies are required to evaluate the model performance across even broader environmental datasets. Future research directions should consider integrating multispectral or thermal imagery, as successfully demonstrated in prior research by Wen et al. [16] and Liu et al. [9]. Such integrations have shown promise in enhancing detection reliability under conditions of heavy smoke or low visibility. Additionally, leveraging edge-computing advances, as discussed in recent studies [17,18], could further reduce latency, improving response times during critical fire events. Collaborative approaches incorporating Internet-of-things (IoT) frameworks [12] may also offer promising avenues for integrating multiple data sources to achieve even greater detection accuracy and reliability in agricultural environments. We acknowledge a critical limitation of the current study regarding practical and long-term validation. Specifically, the dataset used in this study, while robust and diverse, lacks longitudinal data across multiple years and extensive geographical areas. Thus, statistical comparisons of fire-detection response times and model validity over at least five years across different methods were not feasible within this research scope. To address this limitation comprehensively, future studies should undertake systematic data collection and evaluation, incorporating extensive temporal datasets covering larger geographical areas. This approach will enable rigorous statistical analyses, further reinforcing the practical applicability and comparative superiority of our proposed method.

The proposed model significantly advances agricultural fire-detection capabilities, offering a practical, adaptable solution suitable for real-time, resource-limited applications. Its development represents a step forward in environmental monitoring, potentially reducing economic and ecological impacts globally.

6. Conclusions

This study has demonstrated a significant stride forward in the realm of fire detection within agricultural landscapes, addressing urgent needs for timely and accurate fire management. By integrating the SSD with MobileNetV2, we have developed a model that not only surpasses traditional fire-detection systems in speed and accuracy but also adapts seamlessly to the constraints of mobile and edge-computing environments typical of UAVs. Our results indicate that the proposed model achieves a mean Average Precision (mAP) of 97.7%, a clear indication of its superior capability to detect and classify smoke and fire more effectively than existing models. The use of MobileNetV2 as the backbone enhances the model efficiency, enabling it to operate within the limited computational capacities of UAVs without sacrificing performance. This is critical for implementing real-time surveillance and intervention over extensive and often inaccessible agricultural areas. The model high degree of adaptability and its robust performance under varied operational conditions suggest its potential as a transformative tool for precision agriculture. Its deployment could lead to significant reductions in the time between fire onset and detection, thereby minimizing damage and preserving both human life and property. However, challenges such as variability in data quality, environmental conditions, and the need for extensive training datasets remain. Addressing these issues will be crucial for further refining the technology and ensuring its applicability across global agricultural settings.

Future research should explore the integration of additional sensory data, such as thermal and multispectral imagery, to enhance detection capabilities in low visibility conditions caused by smoke or fog. Moreover, leveraging advancements in machine learning and edge computing could facilitate more autonomous operations, making fire-detection systems more resilient and responsive to emergent situations. The successful implementation of this advanced detection system marks a pivotal advancement in agricultural technology, offering a promising avenue for enhancing fire-management practices and contributing to the sustainability and safety of farming operations worldwide. The potential for scaling this technology to other areas of disaster management and environmental monitoring holds great promise for future research and application.

Author Contributions: Conceptualization, S.U. and A.M.; Methodology, A.A., S.U., A.M. and Y.I.C. Software, A.A. and A.M.; Validation, A.A.; Formal analysis, A.A., K.T., A.M. and I.A.; Investigation, K.T., G.B. and I.A.; Resources, K.T., N.E., G.B., Z.T., Y.I.C. and I.A.; Data curation, N.E., G.B., Z.T. and I.A.; Writing—original draft, A.A. and Y.I.C. Writing—review & editing, A.A., S.U. and Y.I.C. Visualization, N.E. and Z.T.; Supervision, S.U., Z.T. and Y.I.C. Project administration, S.U. and Y.I.C. Funding acquisition, Y.I.C. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by Korea Agency for Technology and Standards in 2022, project number is 1415181629.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

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

References

- Morchid, A.; Oughannou, Z.; El Alami, R.; Qjidaa, H.; Jamil, M.O.; Khalid, H.M. Integrated internet of things (IoT) solutions for early fire detection in smart agriculture. *Results Eng.* **2024**, *24*, 103392. [[CrossRef](#)]
- Maity, T.; Bhawani, A.N.; Samanta, J.; Saha, P.; Majumdar, S.; Srivastava, G. MLSFDD: Machine Learning-Based Smart Fire Detection Device for Precision Agriculture. *IEEE Sens. J.* **2025**, *25*, 8921–8928. [[CrossRef](#)]
- Li, L.; Awada, T.; Shi, Y.; Jin, V.L.; Kaiser, M. Global Greenhouse Gas Emissions from Agriculture: Pathways to Sustainable Reductions. *Glob. Change Biol.* **2025**, *31*, e70015. [[CrossRef](#)] [[PubMed](#)]
- Makhmudov, F.; Umirzakova, S.; Kutlimuratov, A.; Abdusalomov, A.; Cho, Y.-I. Advanced Object Detection for Maritime Fire Safety. *Fire*. **2024**, *7*, 430. [[CrossRef](#)]
- Patrick, M.; Mass, C. Weather Associated with Rapid-Growth California Wildfires. *Weather. Forecast.* **2025**, *40*, 347–366.
- Yang, S.; Huang, Q.; Yu, M. Advancements in remote sensing for active fire detection: A review of datasets and methods. *Sci. Total Environ.* **2024**, *943*, 173273. [[CrossRef](#)] [[PubMed](#)]
- Ramos, L.; Casas, E.; Bendek, E.; Romero, C.; Rivas-Echeverría, F. Hyperparameter optimization of YOLOv8 for smoke and wildfire detection: Implications for agricultural and environmental safety. *Artif. Intell. Agric.* **2024**, *12*, 109–126. [[CrossRef](#)]
- Duong, O.; Crew, J.; Rea, J.; Haghani, S.; Striki, M. A Multi-Functional Drone for Agriculture Maintenance and Monitoring in Small-Scale Farming. In Proceedings of the 2024 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 6–8 January 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–4.
- Xie, X.; Zhang, Y.; Liang, R.; Chen, W.; Zhang, P.; Wang, X.; Zhou, Y.; Cheng, Y.; Liu, J. Wintertime Heavy Haze Episodes in Northeast China Driven by Agricultural Fire Emissions. *Environ. Sci. Technol. Lett.* **2024**, *11*, 150–157. [[CrossRef](#)]
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14. Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
- Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.

12. Sharma, A.; Singh, K. UAV-based framework for effective data analysis of forest fire detection using 5G networks: An effective approach towards smart cities solutions. *Int. J. Commun. Syst.* **2025**, *38*, e4826.
13. Yan, X.; Wang, W.; Lu, F.; Fan, H.; Wu, B.; Yu, J. GFRF R-CNN: Object Detection Algorithm for Transmission Lines. *Comput. Mater. Contin.* **2025**, *82*, 1439–1458.
14. Cheknane, M.; Bendouma, T.; Boudouh, S.S. Advancing fire detection: Two-stage deep learning with hybrid feature extraction using faster R-CNN approach. *Signal Image Video Process.* **2024**, *18*, 1–8.
15. Wang, Z.; Xu, L.; Chen, Z. FFD-YOLO: A modified YOLOv8 architecture for forest fire detection. *Signal Image Video Process.* **2025**, *19*, 265. [[CrossRef](#)]
16. Wen, H.; Hu, X.; Zhong, P. Detecting rice straw burning based on infrared and visible information fusion with UAV remote sensing. *Comput. Electron. Agric.* **2024**, *222*, 109078. [[CrossRef](#)]
17. Verma, P.; Bakthula, R. Empowering fire and smoke detection in smart monitoring through deep learning fusion. *Int. J. Inf. Technol.* **2024**, *16*, 345–352. [[CrossRef](#)]
18. Yan, X.; Chen, R. Application Strategy of Unmanned Aerial Vehicle Swarms in Forest Fire Detection Based on the Fusion of Particle Swarm Optimization and Artificial Bee Colony Algorithm. *Appl. Sci.* **2024**, *14*, 4937. [[CrossRef](#)]
19. Gupta, H.; Nihalani, N. An efficient fire detection system based on deep neural network for real-time applications. *Signal Image Video Process.* **2024**, *18*, 6251–6264. [[CrossRef](#)]
20. 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](#)]
21. Mamadaliev, D.; Touko, 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. [[CrossRef](#)]
22. Abdusalomov, A.; Umirzakova, S.; Bakhtiyor Shukhratovich, M.; Mukhiddinov, M.; Kakhorov, A.; Buriboev, A.; Jeon, H.S. Drone-Based Wildfire Detection with Multi-Sensor Integration. *Remote Sens.* **2024**, *16*, 4651. [[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.