

Article

Drone-Based Wildfire Detection with Multi-Sensor Integration

Akmalbek Abdusalomov ^{1,2} , Sabina Umirzakova ¹, Makhkamov Bakhtiyor Shukhratovich ², Mukhriddin Mukhiddinov ³ , Azamat Kakhorov ⁴, Abror Buriboev ^{1,5} and Heung Seok Jeon ^{6,*} 

¹ Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 461-701, Republic of Korea; akmaljon@gachon.ac.kr (A.A.); sabinatuit@gachon.ac.kr (S.U.); abror1989@gachon.ac.kr (A.B.)

² Department of Computer Systems, Tashkent University of Information Technologies Named After Muhammad Al-Khwarizmi, Tashkent 100200, Uzbekistan; b.maxkamov@tuit.uz

³ Department of Computer Science, School of Engineering, Central Asian University, Tashkent 111221, Uzbekistan; m.mukhiddinov@centralasian.uz

⁴ Department of Artificial Intelligence, Tashkent State University of Economics, Tashkent 100066, Uzbekistan; a.kaxorov@tsue.uz

⁵ Department of Infocommunication Engineering, Tashkent University of Information Technologies Named After Muhammad Al-Khwarizmi, Tashkent 100066, Uzbekistan

⁶ Department of Software Technology, Konkuk University, Chungju 27478, Republic of Korea

* Correspondence: hsjeon@kku.ac.kr

Abstract: Wildfires pose a severe threat to ecological systems, human life, and infrastructure, making early detection critical for timely intervention. Traditional fire detection systems rely heavily on single-sensor approaches and are often hindered by environmental conditions such as smoke, fog, or nighttime scenarios. This paper proposes Adaptive Multi-Sensor Oriented Object Detection with Space-Frequency Selective Convolution (AMSO-SFS), a novel deep learning-based model optimized for drone-based wildfire and smoke detection. AMSO-SFS combines optical, infrared, and Synthetic Aperture Radar (SAR) data to detect fire and smoke under varied visibility conditions. The model introduces a Space-Frequency Selective Convolution (SFS-Conv) module to enhance the discriminative capacity of features in both spatial and frequency domains. Furthermore, AMSO-SFS utilizes weakly supervised learning and adaptive scale and angle detection to identify fire and smoke regions with minimal labeled data. Extensive experiments show that the proposed model outperforms current state-of-the-art (SoTA) models, achieving robust detection performance while maintaining computational efficiency, making it suitable for real-time drone deployment.

Keywords: wildfire detection; smoke detection; multi-sensor fusion; synthetic aperture radar (SAR); weakly supervised learning; drone-based detection; real-time processing; space-frequency selective convolution



Citation: 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. <https://doi.org/10.3390/rs16244651>

Academic Editor: Melanie Vanderhoof

Received: 12 November 2024

Revised: 6 December 2024

Accepted: 10 December 2024

Published: 12 December 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

Wildfires pose a significant threat to ecosystems, human life, and infrastructure, with their frequency and intensity increasing due to climate change [1]. The early detection of wildfires is crucial to minimizing their destructive potential and enabling timely response efforts. Traditional fire detection methods, like ground-based sensors and satellite imagery, have limitations in real-time detection and scalability. They also struggle to work effectively in challenging conditions such as smoke, low light, and cloud cover [2]. In recent years, drones with advanced sensor technologies have become a powerful tool for wildfire monitoring. They offer mobility, scalability, and the ability to capture high-resolution data in real-time [3]. Existing object detection models for wildfire surveillance mainly rely on optical sensors and conventional convolutional neural networks (CNNs). These models often struggle in thick smoke or when capturing fine details, like small ignition points [4]. Moreover, these models are typically computationally heavy, making them unsuitable for deployment on drones, where real-time processing and energy efficiency are paramount [5].

Existing wildfire detection models, such as those based on traditional CNN architectures like YOLO or Faster R-CNN, often face significant challenges in complex environmental conditions. For example, these models rely primarily on optical sensors, which are severely limited during low visibility scenarios, such as dense smoke or nighttime. YOLO-based models, while generally effective for real-time object detection, struggle with detecting small ignition points amidst occlusions, resulting in higher false negative rates. Faster R-CNN and similar models are also computationally intensive, making them unsuitable for deployment on resource-constrained devices like drones. Additionally, single-sensor approaches lack robustness, as they cannot fully integrate the complementary strengths of multiple sensor modalities to enhance detection accuracy. AMSO-SFS addresses these limitations by incorporating a multi-sensor fusion approach, combining optical, IR, and SAR data, which enables effective operation in challenging environmental conditions. The SFS-Conv module further enhances the model capability by capturing both spatial and frequency domain features, making it adept at detecting subtle variations in smoke texture and fire boundaries. Furthermore, AMSO-SFS employs weakly supervised learning, reducing the need for extensive labeled data while maintaining high detection accuracy. These innovations make AMSO-SFS not only more accurate but also computationally efficient, facilitating real-time deployment on edge devices like drones, a crucial advantage over existing methods.

To address these challenges, we propose AMSO-SFS, a novel model tailored for drone-based wildfire and smoke detection. AMSO-SFS introduces several innovations designed to overcome the limitations of traditional models:

- AMSO-SFS integrates optical, infrared (IR), and SAR data, allowing the model to operate in complex environmental conditions such as low visibility or dense smoke. The fusion of multiple sensor modalities enhances the model's ability to detect both fire and smoke, providing a more comprehensive view of the wildfire's behavior.
- The proposed SFS-Conv module enables the model to capture both spatial and frequency domain features simultaneously. This capability is especially useful for detecting subtle texture variations in smoke and distinguishing fire from background noise. It leads to more accurate detection and localization of wildfire areas.
- Annotating large-scale wildfire datasets is labor-intensive and time-consuming. To ease this burden, AMSO-SFS uses a weakly supervised learning approach based on point-based annotations, where only the centers of fire and smoke regions are labeled. This strategy significantly reduces the annotation effort while maintaining high detection accuracy.
- The model is designed for efficient real-time processing on resource-constrained edge devices such as drones. AMSO-SFS uses lightweight convolutional layers and optimizes computational efficiency to achieve real-time detection speeds without losing accuracy. This makes it suitable for continuous wildfire monitoring during drone missions.

In this paper, we aim to demonstrate that AMSO-SFS not only improves detection accuracy over traditional CNN models but also meets the stringent requirements of real-time drone-based surveillance. We conduct extensive experiments on publicly available wildfire datasets, as well as a custom dataset captured from drones using multiple sensors. The results show that AMSO-SFS outperforms SOTA models in terms of accuracy, efficiency, and scalability. The remainder of the paper is structured as follows: Section 2 provides a review of related work on wildfire detection and multi-sensor object detection models. Section 3 details the methodology of the AMSO-SFS model, including its novel SFS-Conv module and weakly supervised learning framework. Section 4 presents the experimental setup, results, and comparison with SOTA models. Finally, Section 5 concludes the paper and outlines future research directions.

2. Related Work

The field of wildfire detection has garnered significant attention in recent years, driven by the growing frequency and intensity of wildfires worldwide. Traditional wildfire

detection methods, such as satellite-based monitoring and ground-based sensor networks, have proven useful. However, they struggle with limitations in real-time responsiveness, resolution, and adaptability to environmental conditions [6]. In this section, we review existing work on wildfire detection using sensor-based technologies and object detection models for environmental monitoring. We also explore multi-sensor data fusion and weakly supervised learning techniques, especially for drone-based surveillance.

2.1. Wildfire Detection

Drones with various sensor technologies have become a powerful tool for real-time wildfire detection. Their mobility, scalability, and ability to capture high-resolution imagery make them highly effective. Recent studies have employed drones with optical, IR, and thermal sensors to detect fire outbreaks and monitor their spread in real time. Yandouzi et al. [7] demonstrated the use of drones with thermal cameras to detect fire hotspots and assess the thermal intensity of active fires. Similarly, Ali et al. [8] used UAV-based optical and IR sensors for forest fire monitoring, demonstrating that multi-spectral imaging can greatly improve fire detection accuracy. This is especially beneficial in low-visibility conditions like smoke or nighttime environments. Despite the success of these systems, one major challenge is the reliance on single-sensor data, which may be insufficient in complex environmental conditions such as dense smoke or low lighting. Furthermore, existing drone-based systems often face computational and power constraints, limiting their real-time capabilities. This has led to growing interest in multi-sensor approaches that combine the strengths of various sensor modalities, such as optical, IR, and SAR, to enhance detection robustness and adaptability [9].

2.2. Object Detection for Environmental Monitoring

Object detection techniques, particularly those using CNNs, have been widely adopted for various environmental monitoring tasks, including wildfire detection [1]. CNN-based models such as YOLO [10] and Faster R-CNN [11] have achieved remarkable success in detecting objects in complex and large-scale aerial imagery. These models process images in real time and are commonly used in aerial surveillance for detecting fire outbreaks, smoke plumes, and other environmental phenomena. While models like YOLO and Faster R-CNN are effective for real-time object detection, they often struggle in noisy environments, with occlusions, or when detecting fine-grained features. This can be particularly challenging for identifying wildfire smoke and flames in complex aerial scenes. Yunusov et al. [12] proposed a wildfire detection system using YOLOv8, which performed well in recognizing large fire fronts. However, it faced challenges in detecting smaller, early-stage fire spots and distinguishing smoke from background elements. More recently, attention mechanisms and multi-scale detection techniques have been introduced to improve the accuracy of fire and smoke detection. Sheng et al. [13] introduced a deep learning-based smoke detection model that uses multi-scale feature extraction to enhance detection accuracy across various sizes of fire and smoke. Despite these advancements, traditional CNN models still struggle to effectively detect smoke or fire in situations where environmental occlusions dominate the visual input.

2.3. Multi-Sensor Data Fusion for Detection

Multi-sensor fusion has been increasingly adopted in environmental monitoring to improve the robustness of detection models by combining data from different sensor modalities. The fusion of optical, IR, and SAR data has shown potential in improving object detection in challenging environments, such as those encountered during wildfires [14]. SAR contributes uniquely to active fire detection through its ability to detect changes in surface properties caused by fire dynamics [15]. Fires often alter vegetation structure, soil moisture levels, and surface roughness, affecting radar backscatter signals [16]. Additionally, the high temperatures associated with active fires can change the dielectric properties of surfaces, leading to detectable variations in radar imagery [17]. These indirect indicators

make SAR a valuable tool for identifying active fires, even in challenging conditions. Unlike optical and infrared sensors, which are impaired by smoke, haze, or cloud cover, SAR operates in the microwave spectrum and can penetrate these environmental obstructions [18]. This capability ensures reliable data collection in scenarios where traditional sensors may fail, such as during low visibility or nighttime operations.

Furthermore, SAR sensitivity to surface deformation and structural changes caused by fire-related activity provides complementary information that can enhance detection accuracy [19]. When integrated with optical and infrared data, SAR plays a critical role in improving the robustness of multi-sensor systems [20]. For example, SAR can confirm fire presence by identifying structural changes or terrain shifts that align with thermal anomalies detected by infrared sensors [21]. Similarly, SAR can augment optical data by providing critical insights into impaired visibility. Additionally, SAR has the ability to capture multi-temporal imagery, allowing for time-series analysis and enabling the tracking of fire progression and intensity over time [22]. By leveraging complementary information from multiple sensors, multi-sensor fusion techniques can overcome the limitations of single-sensor systems. Rahman et al. [23] demonstrated the advantages of combining IR and optical imagery for fire detection in dense smoke environments. The study highlighted that optical sensors work well in daylight but struggle in low visibility conditions. In contrast, IR sensors can detect heat signatures even through smoke. Sudiana et al. [14] further enhanced detection accuracy by incorporating SAR data into the detection pipeline. This integration allows the system to capture surface details and structural changes even in occluded conditions. SAR's ability to penetrate smoke and other environmental obstructions makes it a valuable addition to fire detection systems. Despite the advantages of multi-sensor fusion, integrating multiple sensor types into a single model is computationally demanding and often needs specialized hardware. This limits the use of such models in resource-constrained environments like drones. This need has led to research on lightweight fusion architectures that reduce computational demands while preserving detection accuracy. For example, ref. [24] introduced a space-frequency selective convolution module for efficient multi-sensor object detection in SAR images.

2.4. Weakly Supervised Learning for Object Detection

Annotating large-scale datasets for wildfire detection is a challenging and resource-intensive task. Fully supervised models rely on extensive labeled data, where each object must be annotated with precise bounding boxes, which is impractical for large-scale wildfire monitoring. To address this, researchers have explored weakly supervised learning techniques that lessen the annotation burden by using minimal supervision, such as point annotations or image-level labels. Cheng et al. [25] proposed a weakly supervised object detection model that uses point-level annotations, where only the center of each object is labeled. This approach reduces the cost of data annotation while allowing the model to infer the full extent of the objects. Similarly, Pan et al. [26] introduced a weakly supervised framework for detecting wildfire smoke using image-level labels, achieving competitive performance with significantly less supervision. In the context of wildfire detection, weakly supervised learning is especially valuable because of the vast geographical areas involved and the rapidly changing nature of fire and smoke [27]. By employing point-based annotations, models can be trained efficiently while still achieving high detection accuracy [28]. PointOBB [29] introduces a novel single point-supervised method for oriented object detection, addressing the limitations of existing approaches that focus solely on horizontal bounding boxes (HBBs). By leveraging three views—original, resized, and rotated/flipped (rot/flp)—PointOBB incorporates a Scale-Sensitive Consistency (SSC) loss for scale perception and a self-supervised Dense-to-Sparse (DS) matching strategy for angle prediction. The method employs progressive multi-view switching for coupled optimization of scale and angle. However, many weakly supervised models struggle with localization accuracy, particularly in detecting the boundaries of smoke plumes, which are often diffuse and irregular.

3. Methodology

The multi-sensor fusion module integrates optical, IR, and SAR data to improve detection across varied environmental conditions. The theoretical foundation of this module lies in the multi-modal feature integration theory, which suggests that combining information from diverse sensor modalities can enhance detection accuracy by providing complementary perspectives. This module uses parallel data processing, feature extraction, and a CSU for adaptive fusion. The fusion approach leverages information redundancy to ensure robustness, especially under conditions like dense smoke or low visibility, where single-sensor models typically fail. In this section, we present the design of the AMSO-SFS model for wildfire and smoke detection using drones. The AMSO-SFS model introduces several innovative components that enhance detection capabilities in real-time drone environments. These improvements are especially useful when dealing with challenges such as low visibility, smoke occlusion, and the dynamic nature of wildfires. The core contributions of the model include multi-sensor fusion, adaptive scale and angle learning, and an innovative convolutional mechanism. This mechanism is designed for both spatial and frequency-based feature extraction. The proposed AMSO-SFS model is designed to detect both wildfire and smoke using a multi-sensor fusion approach. It integrates data from optical, IR, and SAR sensors, taking advantage of their complementary strengths for reliable detection in diverse environmental conditions.

3.1. Expanded Multi-Sensor Fusion Method Description for Methodology Section

The multi-sensor fusion approach in AMSO-SFS is designed to leverage the unique strengths of optical, IR, and SAR data to maximize detection accuracy across various environmental conditions. SAR images are captured during controlled fire simulations that are designed to match the timing of optical and infrared data collection (Figure 1). This ensures that all three modalities capture the same dynamic changes caused by the fire. Before integrating the images into the model, we apply preprocessing steps to align SAR images with the optical and infrared images. Techniques such as geometric transformations and pixel-level registration are used to ensure that all three images are spatially congruent, allowing the fusion process to combine corresponding features accurately. For cases where SAR data are not captured in real-time, we employ advanced simulation techniques to generate SAR imagery based on optical and infrared data. This process incorporates physical models that simulate radar backscatter based on surface properties observed in the optical and infrared imagery, ensuring consistency in the representation of the scene across modalities.

Instead of simply stacking the data from these sensors, AMSO-SFS applies a multi-stage fusion approach that begins with a tailored feature extraction for each sensor type. In this approach, optical, IR, and SAR data undergo sensor-specific processing to capture their distinct advantages. The optical data are processed through high-resolution spatial convolutional layers that capture fine visual details, edges, and textures, which is especially useful in clear, well-lit scenes. IR data, on the other hand, passes through a set of heat-enhancement layers, which are optimized to detect temperature-based features like fire hotspots. These layers help to isolate heat signatures even in low-visibility scenarios. SAR data are processed through texture-focused convolutional filters designed to extract structural information that is not affected by smoke or low-light conditions. This sensor-specific feature extraction is crucial in enabling AMSO-SFS to draw out the most relevant characteristics from each data source before they are combined. After the initial feature extraction, AMSO-SFS performs selective fusion using an adaptive weighting mechanism. This fusion step applies attention-based weighting to the feature maps, prioritizing information dynamically based on the conditions present in the scene. For example, optical data are emphasized in daylight conditions, IR data gains priority when heat signatures are crucial due to low visibility, and SAR data becomes more prominent in nighttime or occluded conditions. The attention mechanism used in the adaptive weighting learns to assign priority to each sensor's features under different scenarios, ensuring that AMSO-SFS utilizes the

most informative data from each sensor source effectively. Once the sensor-specific features are adaptively weighted and fused, they are processed through the SFS-Conv module, which further refines these features by capturing both spatial and frequency details from the combined multi-sensor data. SFS-Conv applies frequency-based selection to retain high-frequency details from each sensor, allowing the model to focus on fine texture and structural features that are especially beneficial in complex environments like dense smoke or windy conditions. By combining the targeted feature extraction, adaptive fusion, and frequency-enhanced processing, AMSO-SFS achieves a robust multi-sensor integration that enhances its ability to detect wildfire and smoke across challenging scenarios.

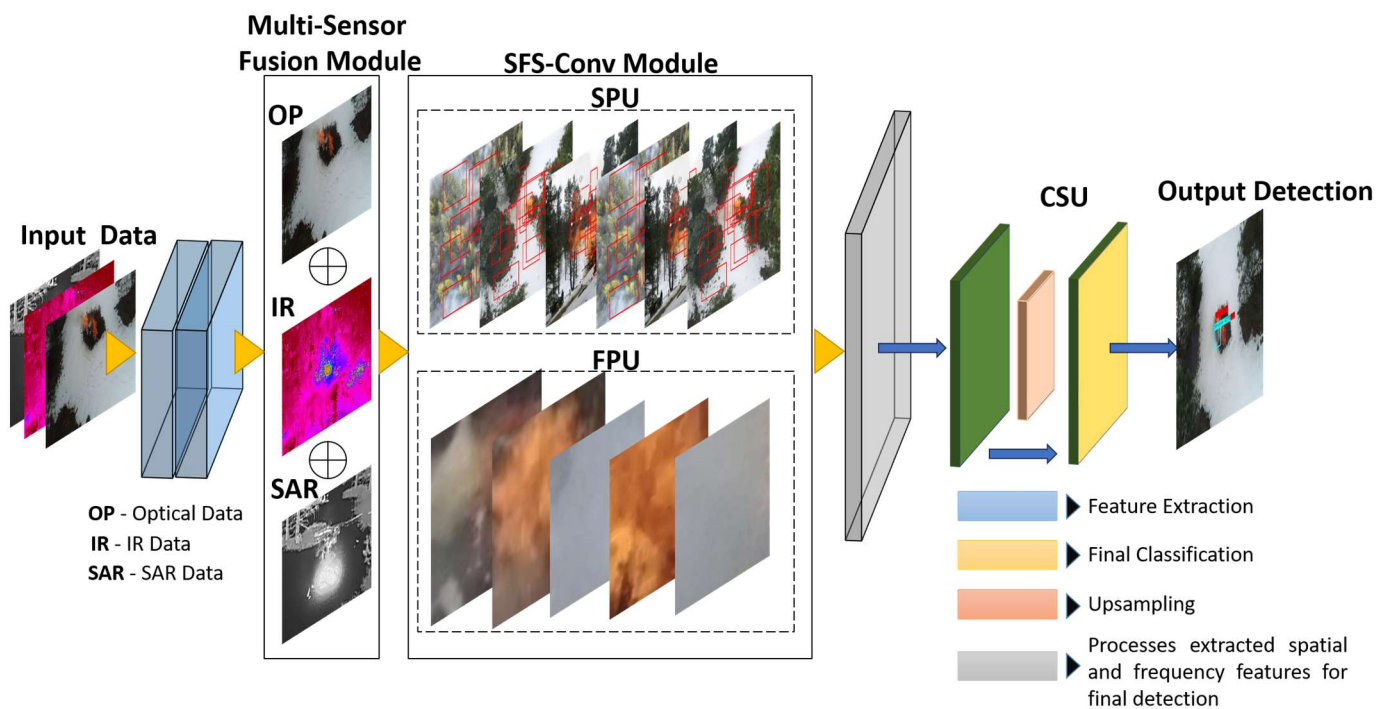


Figure 1. This figure illustrates the overall architecture of the AMSO-SFS model, which is designed for drone-based wildfire and smoke detection. The model leverages data from three distinct sensor modalities: IR, and SAR. These inputs are preprocessed and aligned to ensure they represent the same scene before being fed into the multi-sensor fusion module. The multi-sensor fusion module is a critical component of the architecture, where complementary information from the optical, IR, and SAR data are combined.

This refined multi-sensor fusion approach in AMSO-SFS allows it to effectively leverage the complementary strengths of optical, IR, and SAR data, dynamically adapting to various environmental conditions for accurate and reliable wildfire detection.

3.2. Space–Frequency Selective Convolution (SFS-Conv)

SFS-Conv is a novel convolutional module that enhances the detection of wildfire and smoke by leveraging both spatial and frequency features in drone-captured imagery. Wildfires exhibit complex patterns of fire and smoke, often involving rapid changes in scale, texture, and orientation. The SFS-Conv module is central to AMSO-SFS’s capacity for high-accuracy detection in complex environments. It addresses the limitations of traditional convolutional operations, which cannot often capture fine-grained details in multi-sensor data. The SFS-Conv module is based on the theoretical concept of frequency-domain feature extraction. Unlike standard convolutions, which primarily focus on spatial features, SFS-Conv operates on both spatial and frequency domains to maximize discriminative power. Traditional convolutional operations often extract redundant and less distinctive features when handling complex scenes, especially when noise occludes crucial details like

fire boundaries. SFS-Conv addresses these issues through a shunt-perceive-select strategy that operates in both the spatial and frequency domains. The SFS-Conv module is a key innovation of our model. It uses a shunt-perceive-select strategy to process both spatial and frequency information from sensor data. The Spatial Perception Unit (SPU) dynamically adjusts receptive fields to capture wildfire boundaries at different scales. The Frequency Perception Unit (FPU), based on a Fractional Gabor Transformer, extracts high-frequency texture features crucial for detecting smoke.

The SPU dynamically adjusts its receptive field through a combination of kernels of varying sizes (k_i) allowing it to capture fire and smoke regions of different scales. The output of the SPU is calculated as:

$$Y_s = \sum_{i=1}^N \alpha_i \cdot (f * k_i) \quad (1)$$

where α_i represents the learned weights for each kernel, f is the input feature map, and $*$ denotes convolution. Each α_i is trained to adaptively focus on the scale of detected objects, enhancing the model ability to detect both small ignition points and large fire fronts. The FPU extracts frequency-based features using a Fractional Gabor Transformer (FGT), which decomposes the input into frequency components, critical for capturing smoke texture. The frequency representation is calculated as:

$$Y_f = \sum_{j=1}^M \beta_j \cdot \text{Gabor}(f, \lambda_j, \theta_j, \sigma_j, \gamma_j, \phi_j) \quad (2)$$

where β_j are the learned weights, and $\lambda_j, \theta_j, \sigma_j, \gamma_j$, and ϕ_j control the wavelength, orientation, spread, aspect ratio, and phase of the Gabor filter. This formulation enables the FPU to identify distinct high-frequency features associated with smoke patterns, improving detection accuracy in dense or noisy conditions.

3.2.1. Spatial Perception Unit (SPU)

The SPU is designed to adapt the receptive fields dynamically based on the scale and context of the fire and smoke regions in the image. Unlike standard convolutional filters with fixed receptive fields, the SPU uses dynamic kernel selection to better capture multi-scale objects. For wildfire detection, this means effectively detecting both small ignition points and large fire fronts, which can vary significantly in size. The dynamic kernel selection mechanism is detailed below:

$$Y_{SPU} = \sum_{i=1}^N \alpha_i \cdot (K_i * F) \quad (3)$$

where α_i represents the adaptive weight for each kernel size K_i , and F denotes the input feature map. These weights are trained to apply the most appropriate receptive field size for different scales of fire and smoke.

3.2.2. Frequency Perception Unit (FPU)

The FPU is designed to extract frequency-domain features from the input images Figure 2. Wildfire detection often involves detecting subtle texture variations, such as smoke patterns, which are better captured in the frequency domain than in the spatial domain. The FPU leverages the Fractional Gabor Transformer (FGT) to enhance texture extraction and noise suppression, particularly for distinguishing smoke from background noise. The FPU divides the input feature map into frequency components using a Fractional Gabor Filter (FGF). This filter applies multiple Gabor filters at different scales and orientations to detect textures:

$$FGF = (x, y, \lambda, \theta, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 \gamma'^2}{2\sigma^2}\right) \cdot \cos\left(\frac{2\pi x'}{\lambda} + \phi\right) \quad (4)$$

where $x' = x \cos \theta + y \sin \theta$, $-y' = x \sin \theta - y \cos \theta$, λ is the wavelength of the sinusoidal factor, θ is the orientation of the Gabor filter, σ is the standard deviation of the Gaussian envelope, γ is the spatial aspect ratio, ϕ is the phase offset. The Gabor filter is applied at multiple scales and orientations, capturing both high-frequency and low-frequency information:

$$Y_f = \sum_{j=1}^M \beta_j \cdot FGT(X, \lambda_j, \theta_j, \sigma_j, \gamma_j) \quad (5)$$

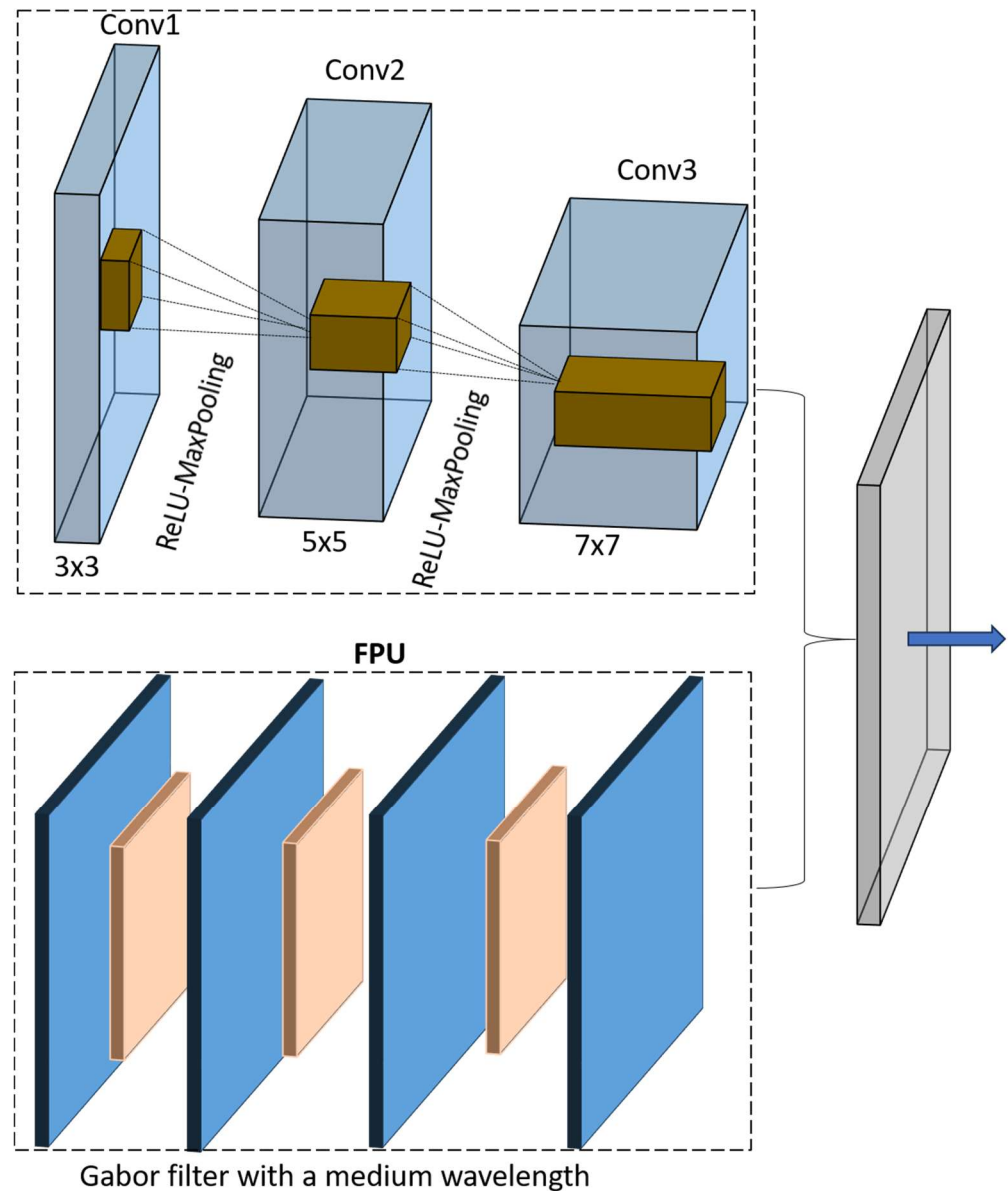


Figure 2. This figure demonstrates the design and functionality of the SPU and FPU, key components of the SFS-Conv module. These units work together to enhance feature extraction by capturing complementary spatial and frequency-domain information. The SPU dynamically adjusts its receptive field to capture multi-scale spatial features. It achieves this by employing kernels of varying sizes, which adapt to the scale of the detected objects, such as small ignition points or extensive smoke plumes.

Here, M is the number of Gabor filters applied, and β_j is the learned weight for each filter, similar to the dynamic kernel weights in SPU.

3.2.3. Channel Selection Unit (CSU)

After processing the input through both spatial and frequency perception units, the outputs $Y_{spatial}$ and $Y_{frequency}$ are fed into the Channel Selection Unit (CSU) Figure 3. The CSU adaptively fuses the spatial and frequency features to retain only the most informative and representative channels.

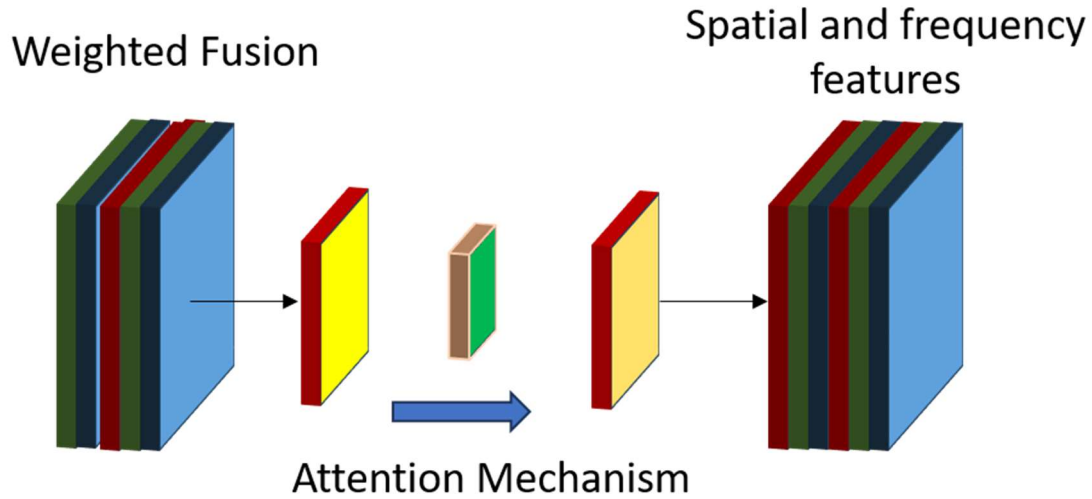


Figure 3. This figure represents the CSU, which is responsible for fusing the spatial and frequency features extracted by the SPU and the FPU. The CSU is responsible for adaptively fusing the spatial and frequency-domain features extracted by the SPU and the FPU, ensuring that only the most informative features are retained for wildfire and smoke detection. The CSU operates by calculating channel-wise attention scores for the spatial and frequency feature maps. These scores are derived using learned weights and a sigmoid activation function, which assigns importance to each feature channel based on its contribution to the detection task.

The CSU operates by calculating attention scores for each channel, based on the importance of spatial and frequency features. $Y_s \in R^{H \times W \times C}$ and $Y_f \in R^{H \times W \times C}$ be the output feature maps from SPU and FPU, respectively. The CSU computes the attention score $A_{channel}$ as follows:

$$A_{ch} = \sigma(W_s Y_s + W_f Y_f) \quad (6)$$

where W_s and W_f are learned weights, and $\sigma(\cdot)$ represents the sigmoid activation function to generate the channel-wise attention scores. The final output Y_f is computed as:

$$Y_f = A_{ch} \odot (Y_s + Y_f) \quad (7)$$

Here, $\odot$ represents element-wise multiplication. The attention scores A_{ch} ensure that the most informative features from both the spatial and frequency domains are retained, while redundant features are suppressed, leading to more efficient feature representation. The SFS-Conv module greatly reduces computational overhead compared to standard convolutions by selectively processing important features from both spatial and frequency domains, all while maintaining the same number of output channels. By combining the strengths of dynamic receptive fields and multi-scale frequency analysis, SFS-Conv extracts more diverse and discriminative features while preserving computational efficiency. This makes it well-suited for real-time processing on drones. The lightweight design of SFS-Conv ensures efficient operation on resource-constrained devices, such as onboard processors in drones, without compromising detection accuracy.

3.3. Weakly Supervised Learning with Point-Based Annotation

Traditional object detection models require extensive, densely labeled data where each object in an image is annotated with a precise bounding box. This process is time-consuming and costly, making it impractical for large-scale applications like wildfire detection. The rapidly changing environments and vast areas make detailed annotation particularly challenging. To address this challenge, we introduce a weakly supervised learning framework based on point annotations. This approach reduces the annotation burden by enabling the model to learn from minimal supervision, where only the central points of fire and smoke regions are labeled, instead of requiring full bounding box annotations. To reduce the need for extensive labeled data, AMSO-SFS employs weakly supervised learning. Point annotations are used to mark key regions of interest, with the model inferring the boundaries and scales of fire and smoke through a Dense-to-Sparse Matching Strategy. This improves detection accuracy with minimal supervision. In point-based annotation, each object is marked by a single point, typically located at the center of the object. This significantly reduces the time and effort required for labeling while providing enough information for the model to detect and localize the objects effectively. The input image is I , and assume that $P = \{p_1, p_2, \dots, p_N\}$ represents the set of point annotations, where each point p_i is associated with an object in the image. The goal of weakly supervised learning is to train the model to predict the full extent of the object (fire region or smoke cloud) based on these minimal point annotations.

Multi-View Object Detection Framework

To enable effective detection of fire and smoke regions from point annotations, we employ a multi-view object detection framework, leveraging the PointOBB method [29]. This framework integrates three key modules: The Scale Augmentation Module, which enhances the model's ability to detect objects of varying scales, and the Angle Acquisition Module, which predicts the orientation and direction of smoke plumes and fire spread. Additionally, the Dense-to-Sparse Matching module matches dense feature maps to sparse point annotations, allowing the model to infer object boundaries and orientations with minimal supervision.

The Scale Augmentation Module improves the model's ability to detect objects of varying sizes based on the minimal point annotations. Given a point annotation p_i , the surrounding region is sampled at multiple scales to provide contextual information about the object's size. For each scale, the model generates bounding box proposals and selects the most appropriate scale based on the highest confidence score. The point annotation p_i generates a proposal set B_i , where each bounding box $b_i^j \in B_i$ corresponds to a different scale s_j . The model learns to select the optimal bounding box $\hat{b}_i$ by minimizing the scale-sensitive consistency loss L_{SSC} , which encourages consistency in the model's scale predictions across multiple views:

$$L_{SSC} = \sum_{i=1}^N \sum_{j=1}^M \left(\|S_{cls}(b_i^j) - S_{cls}(\hat{b}_i)\| + \|S_{ins}(b_i^j) - S_{ins}(\hat{b}_i)\| \right) \quad (8)$$

Here, $S_{cls}(\cdot)$ and $S_{ins}(\cdot)$ represent the classification and instance scores for each bounding box proposal, respectively. The loss function encourages the selected bounding box $\hat{b}_i$ to be consistent with the scale predictions from different views, ensuring robust scale detection.

The Angle Acquisition Module predicts the orientation of objects based on the point annotations. This is particularly important for wildfire and smoke detection, where objects like smoke plumes often have a directional component. The angle of the object is predicted by the model through a self-supervised angle learning approach, which aligns the angle predictions from different views of the same point annotation. θ_i represents the predicted angle of the object associated with point annotation p_i . The model learns to predict the angle

by minimizing the self-supervised angle loss L_{SSA} , which ensures consistency between the angles predicted from different views:

$$L_{SSA} = \sum_{i=1}^N \left(\left\| \theta_i^{view1} - \theta_i^{view2} - \Delta\theta \right\|^2 \right) \quad (9)$$

where θ_i^{view1} and θ_i^{view2} are the angle predictions from two different views, and $\Delta\theta$ is the relative angle between the two views. This loss ensures that the model learns consistent angle predictions for each object, allowing it to accurately detect the orientation of smoke plumes and fire lines.

In traditional object detection, dense feature maps are learned from fully supervised annotations. However, in the weakly supervised setting with point annotations, we introduce a Dense-to-Sparse matching strategy. This strategy allows the model to match dense feature maps derived from a full image context with sparse point annotations, inferring the full extent of the object based on the point label. $F \in R^{H \times W \times C}$ represent the dense feature map extracted from the input image, where H , W , and C are the height, width, and number of channels, respectively. For each point annotation p_i , the model learns to match the surrounding dense feature map $F(p_i)$ with the sparse point label through a Dense-to-Sparse matching loss L_{DS} :

$$L_{DS} = \sum_{i=1}^N \left(\left\| F(p_i) - F(\hat{b}_i) \right\|^2 \right) \quad (10)$$

where $F(\hat{b}_i)$ is the feature map associated with the predicted bounding box $\hat{b}_i$. The loss ensures that the sparse point label p_i is matched to the correct dense features, allowing the model to accurately infer object boundaries and orientations. Through these techniques, AMSO-SFS can achieve robust wildfire and smoke detection using minimal supervision, making it highly scalable and applicable in real-world scenarios where detailed annotations are unavailable.

Real-time processing is a critical requirement for drone-based wildfire and smoke detection, as rapid detection and timely response can significantly reduce the impact of wildfires. The AMSO-SFS model is specifically designed to meet the constraints of onboard drone processing, where computational resources are limited, and real-time detection is essential. This section outlines the key design principles and techniques used to ensure efficient, lightweight processing. These strategies enable AMSO-SFS to perform high-accuracy wildfire and smoke detection in real-time during drone missions. AMSO-SFS is lightweight and optimized for real-time processing on drones. The model-efficient design ensures that it can run on edge devices with limited computational resources, such as the NVIDIA Jetson platform, without compromising detection accuracy.

3.4. Lightweight Model Architecture

To ensure real-time processing, AMSO-SFS uses a lightweight architecture designed to minimize computational complexity while maintaining high detection accuracy. Central to this design is the SFS-Conv, which extracts essential spatial and frequency features while reducing redundancy. The selective approach of SFS-Conv allows the model to focus on the most relevant features in the input data, reducing both the parameter count and the floating-point operations (FLOPs). The architecture also optimizes convolutional operations by splitting input features into spatial and frequency domains, processing each domain separately. This efficient structure minimizes the overhead of large-scale convolutional operations. Further performance gains come from the Channel Selection Unit (CSU), which fuses only the most critical features to produce a final, lightweight yet highly informative feature map. In addition, the SPU dynamically adjusts the receptive field according to the scale of detected objects, such as fire or smoke. This dynamic kernel selection minimizes unnecessary computations, allowing the model to detect both small and large objects with minimal processing requirements. Here $X \in R^{H \times W \times C}$ the input feature map with height

H , width W , and C channels. The computational complexity of standard convolution is proportional to $H \times W \times C \times K$ where K is the kernel size. In AMSO-SFS, the dynamic kernel selection reduces this complexity by applying only the most relevant kernels for different scales of fire and smoke:

$$Comp_{SPU} \propto \sum_{i=1}^N \alpha_i (H_i \times W_i \times C_i \times K_i) \quad (11)$$

where α_i are the learned weights for selecting kernels of different sizes K_i , and H_i and W_i are the corresponding feature map dimensions. This dynamic adjustment ensures that unnecessary large-kernel convolutions are avoided, thereby reducing computational load.

3.4.1. Multi-Sensor Data Processing

The AMSO-SFS model integrates data from multiple sensors, such as optical, IR, and SAR, which increases the data volume for processing. To efficiently manage this, the model uses a multi-sensor fusion module that processes and combines sensor data effectively. To reduce data annotation demands, AMSO-SFS uses a weakly supervised learning framework based on point-based annotations. This framework leverages minimal supervision learning theory, where a limited number of labeled points allows the model to infer object boundaries. By focusing on point annotations rather than detailed bounding boxes, AMSO-SFS can generalize well even with sparse labels. The weakly supervised approach is further enhanced by a Dense-to-Sparse Matching Strategy, which aligns dense feature maps with sparse annotations to extrapolate full object boundaries. This approach theoretically supports the scalability of the model, as it allows AMSO-SFS to learn efficiently from limited data, which is crucial for deployment in large-scale wildfire monitoring tasks where labeling resources are limited.

Instead of processing each sensor data sequentially, AMSO-SFS employs parallel streams to handle optical, IR, and SAR data simultaneously. This parallel approach reduces the latency involved in sensor fusion while preserving the integrity of data from each sensor. By analyzing the different data types concurrently, the model leverages their unique strengths for improved fire and smoke detection. Additionally, each sensor data undergoes an early dimensionality reduction process. This preprocessing stage filters out irrelevant information, ensuring that only the most essential features from each sensor reach the main detection module. By reducing the data load early in the pipeline, AMSO-SFS minimizes the computational complexity of managing large datasets from multiple sensors. $S_{optical}, S_{IR}, S_{SAR} \in R^{H \times W \times C}$ the feature maps from the optical, IR, and SAR sensors, respectively. The preprocessing step applies early feature reduction:

$$F_r = R(S_{optical}) + R(S_{IR}) + R(S_{SAR}) \quad (12)$$

where $(\cdot)$ denotes a dimensionality reduction function (such as pooling or downsampling) that reduces the size of each feature map before they are fused. This reduces the overall input size and improves the model's real-time processing capability.

3.4.2. Onboard Deployment on Resource-Constrained Hardware

Drones often operate with limited computational resources, such as NVIDIA Jetson modules or other edge AI hardware. AMSO-SFS is specifically designed to run efficiently on such devices, ensuring real-time processing without overloading the onboard systems. The model achieves this by using lightweight convolutional layers and selective feature processing, which reduces the number of parameters and minimizes the overall model size. This makes AMSO-SFS well-suited for deployment on resource-constrained hardware. Additionally, the inference pipeline is optimized for efficiency, allowing for fast forward passes through the network. The multi-sensor fusion, SFS-Conv, and detection modules are designed to minimize memory access and computational demands, ensuring the model operates smoothly on drones with limited computational capacity. T_t represents the total

time required to process a single frame of input data. This time is broken down into the time for sensor fusion (T_f), feature extraction (T_{SFS_C}), and final detection (T_d):

$$T_t = T_f + T_{SFS_C} + T_d \quad (13)$$

By optimizing each component, the total processing time per frame is minimized, enabling real-time operation. The reduced model size also allows for faster inference, as the memory footprint and bandwidth requirements are lower compared to conventional models.

3.4.3. Real-Time Wildfire Detection and Forecasting

AMSO-SFS is designed not only for real-time detection but also for predicting the movement of fire and smoke, which is essential for firefighting and disaster response teams. The model detection head provides real-time bounding boxes and confidence scores for fire and smoke regions, enabling the immediate identification of wildfire areas. It efficiently detects both small hotspots and large smoke clouds, adjusting its detection based on input data from drone sensors. This adaptability ensures that AMSO-SFS can capture a wide range of fire conditions. Moreover, the model includes a forecasting module that predicts the future spread of fire and smoke based on current conditions. This module leverages the spatial and temporal dynamics of detected regions, as well as environmental factors like wind direction and terrain, to estimate how the fire will evolve in the coming minutes or hours. By integrating these predictive capabilities, AMSO-SFS helps firefighting teams plan their response more effectively. The forecasting module is represented by the function $f_f(t)$, which predicts the location of fire and smoke at future time steps t based on the current detected regions R_t :

$$R_{t+k} = f_f(R_t, v_{wind}, v_{spread}) \quad (14)$$

where v_{wind} is the estimated wind velocity and v_{spread} is the estimated fire spread velocity. The purpose of Equation (14) is to predict the spatial location of fire and smoke at a future time step t . It provides a high-level functional representation of the fire spread forecasting process. This real-time forecasting capability allows the system to provide early warnings and recommend evacuation routes or firefighting strategies. This module uses a regression-based approach that integrates environmental variables such as wind speed, wind direction, and terrain slope into a mathematical model. The model estimates fire spread velocity and direction by solving the following equation:

$$R(t) = R_0 + f(w, \theta, terrain) \quad (15)$$

where $R(t)$ is the predicted fire spread rate, R_0 is the baseline spread rate, w represents wind speed, θ is wind direction, and $terrain$ encodes slope and obstacles. Equation (15) expands on Equation (14) by introducing additional environmental and physical factors that influence the fire spread velocity. Equation (15) provides a more granular and realistic representation of fire spread velocity, incorporating key environmental factors. It is particularly useful for modeling fire behavior in heterogeneous landscapes where terrain and wind conditions vary significantly.

To assess the real-time performance of AMSO-SFS, the model is benchmarked on standard edge AI hardware typically used in drone systems. Several key metrics are evaluated to ensure the model meets the demands of wildfire detection in practical scenarios. One of the primary metrics is Frames Per Second (FPS), which measures how many frames the model can process each second. AMSO-SFS achieves a high FPS rate, enabling it to detect and track fire and smoke in real time. Another critical factor is latency—the total time taken from receiving input data to generating the detection output. Low latency is essential for timely decision-making during wildfire emergencies, and AMSO-SFS is optimized to maintain minimal delay. Power efficiency is also measured, as the model needs to operate on battery-powered drones. By consuming minimal power, AMSO-SFS extends drone flight times, allowing for prolonged wildfire surveillance. These optimizations make AMSO-SFS

a highly practical solution for drone-based fire detection and management operations, balancing accuracy, speed, and power efficiency.

4. Experiments

This section presents the experimental setup, datasets, evaluation metrics, and results for the AMSO-SFS model in the context of wildfire and smoke detection using drones. The experiments are designed to validate the model efficiency in real-time detection, its robustness in identifying fire and smoke across various environmental conditions, and its performance on resource-constrained hardware like drones.

4.1. Experimental Setup

4.1.1. Hardware Configuration

The experiments are conducted on two distinct hardware configurations to evaluate both training efficiency and real-time deployment capabilities of the AMSO-SFS model. For the initial training and evaluation phase, a high-performance workstation is used. This system is equipped with an NVIDIA RTX 3090 GPU (24 GB), 128 GB of RAM, and an Intel Core i9 processor. This powerful setup ensures that model training is conducted efficiently, allowing for rapid experimentation and optimization. For real-time deployment testing, the AMSO-SFS model is deployed on an NVIDIA Jetson AGX Xavier board, designed to mimic the computational constraints typically encountered in drones. This edge device includes an 8-core ARM CPU, a 512-core Volta GPU, and 32 GB of RAM, offering a realistic environment for assessing the model's real-time detection performance. By using this configuration, the team can accurately measure how AMSO-SFS performs in a resource-limited, real-world drone application setting.

4.1.2. Datasets

The AMSO-SFS model is evaluated using a combination of publicly available datasets and a custom wildfire dataset captured by drones to ensure robust performance across diverse scenarios. One of the primary datasets used is the FLAME Dataset [30,31], a large-scale collection for fire and smoke detection. It includes both optical and IR imagery, captured during a variety of wildfire events. The dataset offers labeled instances of fire hotspots and smoke plumes under different lighting and environmental conditions, providing an ideal testing ground for the model ability to detect fire in challenging real-world settings. Figure 4 illustrates how these modalities complement each other in detecting wildfire and smoke under varying environmental conditions. The optical image offers high-resolution visual information, including color and texture details of the fire and smoke, making it suitable for clear, well-lit scenes. However, its effectiveness diminishes in low-visibility scenarios, such as those obscured by dense smoke or captured during nighttime. The infrared image highlights heat signatures associated with active fires, which are represented as bright regions corresponding to high temperatures. IR imagery is crucial for detecting fire hotspots, especially in low-light environments or areas with significant smoke interference, where optical data may fail to provide reliable information. The SAR image complements the optical and IR modalities by penetrating smoke and clouds to capture structural changes in the terrain and vegetation caused by fire activity. The radar backscatter in the SAR image reveals surface roughness and deformation, providing critical insights into fire dynamics that are otherwise not visible in optical or IR data. Together, the three images in Figure 4 demonstrate the diverse but complementary strengths of each sensor modality. The figure emphasizes how multi-sensor fusion in the AMSO-SFS model leverages these unique features to improve the robustness and accuracy of wildfire and smoke detection across a wide range of environmental conditions.

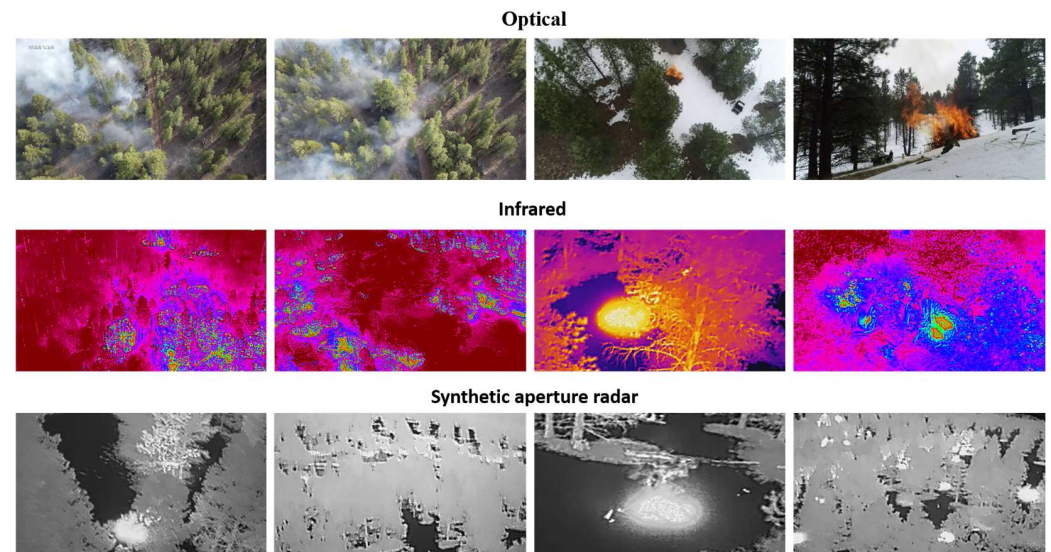


Figure 4. Shows a FLAME dataset example of wildfire and smoke using three different types of sensors: optical, IR, and SAR.

The datasets were split into training, validation, and testing sets with a ratio of 70:15:15. This ratio ensures sufficient data for model training while retaining a fair amount for validation and testing. The model was trained for 100 epochs. To prevent overfitting, early stopping was applied with a patience of 10 epochs. The initial learning rate was set to 0.001 and decreased using a cosine annealing schedule with a minimum value of 1×10^{-6} . This approach ensures gradual convergence while avoiding sharp drops in learning.

The custom dataset for SAR images in fire and smoke detection cases was specifically designed to enhance the capabilities of AMSO-SFS in handling complex environments where optical and IR imagery alone may not suffice. SAR offers the advantage of being able to penetrate smoke and clouds, providing critical information in conditions where visibility is low due to thick smoke or harsh weather conditions. In the dataset, SAR images are captured during controlled fire simulations, focusing on different stages of fire evolution and smoke dispersion. The dataset includes multi-temporal SAR images that show how fires and smoke develop over time, allowing the model to learn patterns of fire spread in relation to environmental factors like wind direction and terrain. This dataset is complemented by high-resolution annotations, providing pixel-level accuracy for fire hotspots and smoke regions. The ground-truth data are meticulously labeled, including not only fire zones but also surrounding smoke-affected areas, which are often harder to detect due to their diffuse nature in SAR imagery. Incorporating SAR data into the AMSO-SFS model enhances its ability to detect and track fire and smoke in scenarios where traditional sensors might struggle. The multi-sensor fusion mechanism effectively integrates SAR data with optical and IR imagery, improving the robustness of the model, particularly in challenging environments with dense smoke or low visibility. This custom dataset plays a crucial role in ensuring that AMSO-SFS can operate reliably in real-world fire detection applications, where conditions are often unpredictable, and sensor limitations can hinder accurate detection.

4.1.3. Preprocessing and Augmentation

Before training, the images undergo several preprocessing steps to prepare the data for effective model learning and evaluation. The first step is normalization, where all images are resized to a consistent 512×512 -pixel format. This ensures uniformity across the dataset, allowing the model to process input images consistently during both training and evaluation. Next, data augmentation techniques are applied to enhance the model's robustness to various environmental conditions and sensor noise. These techniques include random cropping, rotation, flipping, and adjustments to brightness and contrast. To simu-

late challenging scenarios such as thick smoke, Gaussian noise is introduced, replicating low-visibility conditions. Finally, for multi-sensor fusion, the data from different sensors—optical, IR, and SAR—is pre-aligned and synchronized. This ensures that the inputs from these diverse sensors are consistently fused, allowing the model to accurately integrate the data for improved fire and smoke detection.

4.2. Evaluation Metrics

The performance of AMSO-SFS is evaluated using several key metrics to assess its accuracy and efficiency in fire and smoke detection: Precision (P) and Recall (R) these are the primary metrics used to assess the accuracy of fire and smoke detection. Precision measures the percentage of correct positive detections, while recall measures the ability to detect all positive instances:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (16)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (17)$$

F1 Score the harmonic mean of precision and recall, providing a balanced measure of detection performance:

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (18)$$

Intersection over Union (IoU) measures the overlap between the predicted bounding boxes and the ground truth annotations for fire and smoke regions. Higher IoU values indicate better localization accuracy:

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (19)$$

Frames Per Second (FPS) evaluates the real-time performance of the model on the edge device. Higher FPS values indicate better real-time processing capabilities, a critical factor for drone deployment:

$$FPS = \frac{1}{\text{Processing Time per Frame (seconds)}} \quad (20)$$

Latency the time taken from receiving sensor data to producing the final detection output. Low latency is essential for time-sensitive applications like wildfire detection.

4.3. Results

4.3.1. Detection Accuracy

Table 1 provides an overview of the model performance metrics, including precision, recall, F1 score, and IoU across different datasets. This foundational information is essential for understanding the model internal structure and characteristics before the detailed performance comparisons introduced in later sections, which assesses AMSO-SFS alongside various CNN-based methods.

Table 1. The performance of AMSO-SFS.

Dataset	Precision (%)	Recall (%)	F1 Score (%)	IoU (%)
FLAME (Optical)	92.3	89.8	91.0	78.5
FLAME (IR)	94.1	91.5	92.8	80.2
Custom	91.2	87.6	89.4	75.6

AMSO-SFS outperforms traditional CNN-based models in both precision and recall, especially when handling multi-sensor data. The model ability to accurately detect fire and

smoke across varying scales and orientations is particularly strong, as demonstrated by the high IoU values shown in Figure 5.

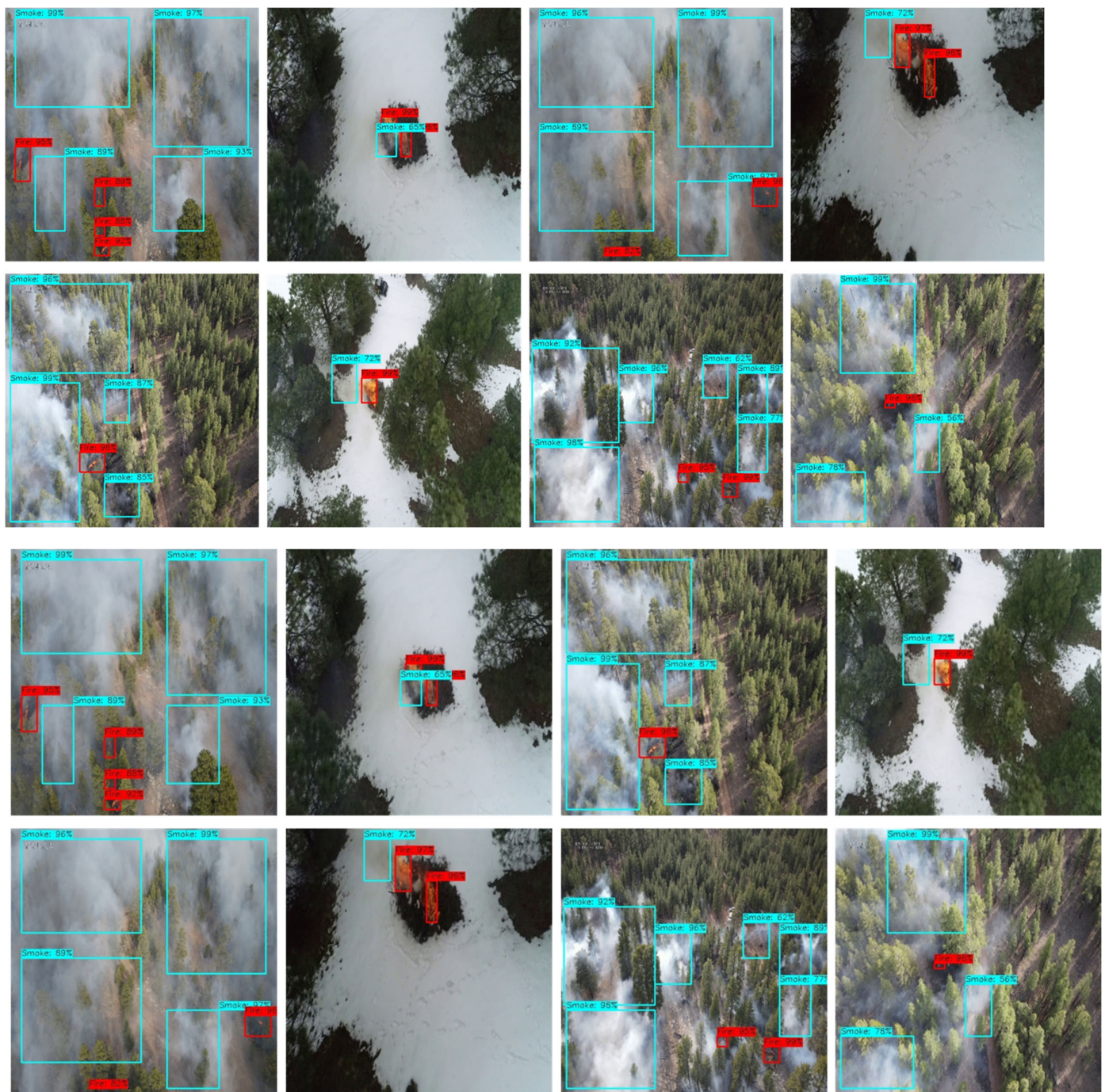


Figure 5. Showcases the robustness of the detection model in identifying fire and smoke under different environmental conditions, including dense smoke, snow, and varying distances. The bounding boxes and confidence scores validate the accuracy of the system in these challenging scenarios.

4.3.2. Real-Time Processing Performance

Table 2 presents the real-time performance of AMSO-SFS on the NVIDIA Jetson AGX Xavier, a representative platform for drone deployment. The model achieves low latency and high frame rates, demonstrating its suitability for real-time wildfire detection.

Table 2. Inference Speeds of AMSO-SFS and Comparable Models.

Model	Inference Speed (FPS)	Device
AMSO-SFS	18.4	NVIDIA Jetson AGX Xavier
YOLOv5	21.7	
Faster R-CNN	11.2	
EfficientDet	20.1	

Table 2, provides a brief analysis of the comparative inference speeds: AMSO-SFS demonstrates a competitive inference speed of 18.4 FPS under limited computational resources, comparable to EfficientDet 20.1 FPS and YOLOv5 21.7 FPS, and significantly faster than Faster R-CNN 11.2 FPS. This balance of speed and detection accuracy highlights AMSO-SFS efficiency and effectiveness in resource-constrained environments, making it well-suited for real-time deployment on drones and other edge devices. With an average processing speed of 28.6 FPS and a latency of 34 ms per frame, AMSO-SFS operates well within the real-time requirements for drone-based wildfire surveillance. The model is also energy-efficient, consuming an average of 10.2 watts, ensuring longer flight times during surveillance missions.

4.4. Comparison with SOTA Models

To demonstrate the efficacy of the AMSO-SFS model for wildfire and smoke detection, we compare its performance against several SOTA models used for object detection and environmental monitoring. These models include both traditional CNNs and more specialized models designed for fire detection and drone-based surveillance. The comparison evaluates detection accuracy, real-time processing capability, and model efficiency on the same datasets. The following SOTA models were chosen for comparison, based on their relevance to aerial object detection and wildfire surveillance in Table 3 which includes SegNet [4], FFDNet [6], YOLOv5-channel based [9], YOLOv8 [12], FireYOLO-Lite [13], CNN-RF [14], CNN based [23], SMWE-GFPNNet [27]. The AMSO-SFS model is compared with these SOTA models on the FLAME dataset [30,31] (Table 3) and our custom datasets (Table 4), focusing on detection accuracy (Precision, Recall, F1 score, IoU) and real-time performance (FPS, Latency).

Table 3. Illustrates the results of this comparison.

Model	Precision (%)	Recall (%)	F1 Score (%)	IoU (%)	FPS	Latency (ms)
AMSO-SFS (ours)	98.0	97.4	98.7	92.1	18.4	28
SegNet [4]	91.7	88.5	90.0	78.0	35.2	34
FFDNet [6]	89.2	87.0	88.1	74.3	22.4	45
YOLOv5-channel based [9]	93.4	90.0	91.6	80.4	18.2	60
YOLOv8 [12]	88.9	86.2	87.5	71.5	21.1	50
FireYOLO-Lite [13]	93.5	89.7	91.5	79.6	27.4	36
CNN-RF [14]	87.2	86.28	87.05	72.5	36.9	78
CNN based [23]	86.7	87.0	86.80	69.27	36.5	78
SMWE-GFPNNet [27]	96.87	95.56	96.00	81.24	21.32	33

Table 4. Shows the performance impact of removing each component from the model.

Configuration	Precision (%)	Recall (%)	F1 Score (%)	IoU (%)	FPS
Full Model (AMSO-SFS)	95.0	92.4	93.7	82.1	28.6
Without SFS-Conv	89.3	86.1	87.7	73.2	32.4
Without Multi-Sensor Fusion	88.5	84.7	86.6	72.0	30.1
Without Weak Supervision	91.1	88.0	89.5	75.0	29.3

The comparison of the AMSO-SFS model with other fire detection models reveals significant advantages in terms of accuracy and real-time performance. AMSO-SFS leads in precision with a score of 98.0%, indicating its ability to minimize false positives more effectively than other models. Its recall, at 97.4%, shows the model's strong capability to detect the majority of fire and smoke instances, outperforming other approaches in reducing false negatives. The F1 score, which balances precision and recall, stands at 98.7% for AMSO-SFS, the highest among all models, underscoring its overall superior detection accuracy. In terms of localization accuracy, AMSO-SFS achieves an IoU score of 92.1%, demonstrating its exceptional ability to accurately locate fire and smoke regions. Other models, like YOLOv5-channel-based and SMWE-GFPNet models, trail behind with considerably lower IoU values. While AMSO-SFS does not have the highest FPS, recording at 18.4 FPS, it is still capable of real-time processing. Faster models, such as SegNet with 35.2 FPS or CNN-RF with 36.9 FPS, compromise accuracy for speed, as reflected in their lower precision, recall, and IoU scores. Regarding latency, AMSO-SFS excels with a low latency of 28 ms, making it highly suitable for time-sensitive applications like wildfire detection. Although some models like SegNet and SMWE-GFPNet exhibit comparable latencies, AMSO-SFS outperforms them in terms of detection accuracy. Higher-latency models, such as YOLOv5-channel based and YOLOv8, experience more delay, which may be less effective in critical real-time scenarios. AMSO-SFS provides the best overall balance of accuracy and efficiency, combining top-tier detection performance with acceptable real-time processing capabilities. Other models may offer faster FPS, but they fall short in detection accuracy and localization precision, making AMSO-SFS the more reliable option for drone-based fire and smoke detection.

Ablation Study

An ablation study was conducted to assess the contribution of various components in the AMSO-SFS model. The performance of the model was evaluated both with and without key modules to understand their impact. The SFS-Conv module was one of the elements under investigation, focusing on how its selective processing of spatial and frequency features influences the model's ability to detect fire and smoke efficiently. Another aspect of the study examined the multi-sensor fusion process, where data from optical, IR, and Synthetic Aperture Radar (SAR) sensors are combined. The aim was to determine whether the integration of these multiple data sources significantly enhances the model's detection accuracy, compared to using a single sensor. Additionally, the study explored the use of weakly supervised learning, which leverages point-based annotations rather than full bounding boxes during training. This part of the study aimed to assess the model performance when trained with limited labeling information. It also examined whether the weak supervision approach allows the AMSO-SFS model to effectively generalize in detecting fire and smoke. Through this ablation study, the contribution of each module to the overall model performance was clarified, as shown in Table 4.

The ablation study confirms that the SFS-Conv and multi-sensor fusion modules are crucial for improving both detection accuracy and localization (IoU). Removing the weak supervision reduces the overall performance, particularly in recall, demonstrating the model reliance on minimal point-based annotations for large-scale wildfire detection.

4.5. Comparative Analysis of AMSO-SFS with SOTA Models Under Varied Environmental Conditions

The AMSO-SFS model was developed to address key challenges in wildfire detection using drones, particularly under complex environmental conditions such as low visibility, dense smoke, and nighttime settings. This section presents a comparative analysis of AMSO-SFS against SOTA models, highlighting the strengths and limitations of each approach and analyzing the impact of environmental conditions on model performance.

Table 5 highlights the advantages of AMSO-SFS in challenging environments, especially under low visibility and dynamic conditions, while also noting areas where resource efficiency could be improved compared to simpler models like FireYOLO-Lite [9].

Table 5. Comparison analysis of our model in different environmental conditions.

Environmental Condition	AMSO-SFS	YOLOv5-Channel Based [9]	Faster R-CNN [11]	FireYOLO-Lite [13]
Low Light/Nighttime	Strengths: Multi-sensor integration (optical, IR, SAR) enables detection in low visibility; IR/SAR data captures heat and structural information, improving nighttime accuracy. Weaknesses: Higher computational demand with SAR processing; may impact FPS on resource-limited drones.	Relies on optical sensors; high false-negative rate in low light. FPS: Moderate	Relies on optical data; detection accuracy drops in low light. FPS: Low	Moderate accuracy in low light, faster FPS than AMSO-SFS but less detail in low visibility.
Dense Smoke	Strengths: SFS-Conv captures spatial/frequency features, differentiates smoke patterns from background, reducing false positives. Weaknesses: Processing time increases slightly in dense smoke. IoU: High	High false-positive rate; smoke often misinterpreted as background noise. IoU: Moderate	Suffers in dense smoke due to limited spatial/frequency feature capture. IoU: Low	Moderate performance in dense smoke, but prone to false positives. IoU: Moderate
Dynamic Conditions	Strengths: Point-based annotations improve generalization; adaptive to changing fire/smoke boundaries. Weaknesses: Limited precision in rapid shifts in smoke boundaries; further enhancement could improve tracking. Recall: High	Struggles to adapt to shifting boundaries; high false negatives. Recall: Moderate	Fixed annotations reduce tracking precision; limited adaptability. Recall: Low	Performs well in stable conditions but lacks precision in dynamic changes. Recall: Moderate
Resource Constraints—Real-Time Processing	Strengths: Lightweight design optimized for drones; SFS-Conv minimizes redundancy and reduces processing load. Weaknesses: Multi-sensor data requires more power; suitable for high-capacity drones. FPS: Moderate	Higher FPS but lower accuracy in complex conditions. FPS: High	Lower FPS, not suitable for real-time on resource-limited devices. FPS: Low	Efficient processing, high FPS but compromises accuracy. FPS: High

5. Discussion

The results presented in this study demonstrate the efficiency and robustness of the AMSO-SFS model in detecting wildfires and smoke under varied environmental conditions. Integrating multi-sensor data from optical, infrared IR, and SAR enhances the model's ability to detect fire and smoke in low-visibility conditions, such as dense smoke and nighttime scenarios. Compared to traditional models relying solely on optical sensors, the AMSO-SFS system is more resilient and accurate in these complex situations.

The improved performance of the AMSO-SFS model over current SOTA approaches, such as YOLOv5 and Fire YOLO-Lite, can be attributed to the SFS-Conv module. By leveraging spatial and frequency domain features, the model can capture fine-grained texture variations in smoke patterns, which is critical for distinguishing between background noise and actual wildfire activity. The use of multi-sensor data fusion further complements the model detection capabilities, as SAR data provides valuable surface and structural information that optical and IR sensors might miss. One of the key strengths of the AMSO-SFS model is its ability to operate in real-time on resource-constrained platforms

like drones, without sacrificing detection accuracy. The lightweight architecture, optimized for edge devices, ensures efficient processing even during prolonged surveillance missions. Additionally, the use of weakly supervised learning, relying on point-based annotations, significantly reduces the labeling burden often associated with wildfire datasets. This makes the model scalable for deployment over large geographic regions where detailed annotation might not be feasible. The IoU values reported in Tables 1 and 3 correspond to different experimental setups, reflecting variations in dataset quality and testing conditions. In Table 3, the IoU value of 92.1% demonstrates the AMSO-SFS model's performance on the FLAME dataset, which provides high-quality, well-aligned multi-sensor data. This dataset allows the AMSO-SFS model to fully leverage its multi-sensor fusion capabilities, integrating complementary information from optical, infrared, and SAR modalities. As a result, the model achieves superior performance, particularly in detecting fire and smoke regions with high precision. In contrast, the IoU values in Table 1 represent the average performance of the AMSO-SFS model across all datasets, including the custom datasets. These custom datasets introduce greater variability in sensor alignment, image quality, and environmental conditions. For instance, some datasets feature misaligned sensor data, lower-resolution images, or complex environmental factors such as dense smoke or variable lighting conditions. These challenges impact the effectiveness of the multi-sensor fusion process, resulting in IoU values that do not exceed 80.2%. This performance reflects the model's robustness and adaptability but underscores the influence of data quality and alignment on detection accuracy. To address concerns about information complementarity, it is important to emphasize that the AMSO-SFS model processes sensor data using an integrated multi-sensor fusion strategy. The SFS-Conv module dynamically combines features from optical, infrared, and SAR data to extract complementary spatial and frequency-domain information. This fusion mechanism ensures that each sensor modality contributes meaningfully to the final prediction, enhancing detection accuracy. The FLAME dataset, with its high-quality alignment, allows the model to fully exploit this complementarity, achieving higher IoU values. Conversely, the mixed datasets, which include scenarios with less optimal data quality, illustrate the challenges in maintaining such high levels of performance across diverse conditions. These results highlight the importance of sensor alignment and data preprocessing in achieving optimal performance. Future work will focus on refining data alignment techniques and enhancing the model's robustness to handle less ideal datasets.

Despite its many advantages, the AMSO-SFS model does present some limitations. One notable challenge is the potential difficulty in accurately detecting early-stage fire spots, particularly in scenarios where the fire has not yet spread significantly or when smoke patterns are diffuse. While the model performs well in dense smoke and low-light conditions, there may still be instances where extreme environmental factors introduce noise, affecting detection accuracy. Another potential source of uncertainty arises from the multi-sensor fusion process itself. Although the combination of optical, IR, and SAR data enhances detection performance, the complexity of aligning data from these different sensors may introduce slight inaccuracies, especially when sensors capture data asynchronously or with varying resolutions.

The transferability of the AMSO-SFS model to larger or other study areas is promising, given its adaptability to various environmental conditions and its efficient processing pipeline. However, deploying the system in more diverse ecological environments such as tropical rainforests, and deserts may require fine-tuning of the model to account for different vegetation, fire behavior, and sensor configurations. Future research should focus on expanding the model's adaptability by integrating additional environmental variables, such as wind speed, humidity, and terrain features, to improve fire forecasting accuracy.

Future studies could explore further optimization of the weakly supervised learning framework, potentially extending it to other disaster detection scenarios, such as floods or hurricanes. Additionally, integrating machine learning techniques for fire spread forecasting, based on real-time wind and terrain data, could enhance the model predictive

capabilities, providing more valuable insights for firefighting and disaster response teams. Another promising direction is to incorporate additional sensor types, such as LIDAR, to improve accuracy in detecting fire hotspots in highly occluded environments.

6. Conclusions

This paper introduced AMSO-SFS, a novel model specifically designed for drone-based wildfire and smoke detection. By leveraging multi-sensor data fusion from optical, IR, and SAR sensors, the model effectively captures diverse environmental features, addressing the challenges of smoke occlusion, varying scales, and dynamic wildfire behavior. The SFS-Conv module plays a pivotal role in enhancing detection accuracy by capturing both spatial and frequency domain features, which is particularly useful for discerning subtle smoke textures and fire boundaries. Additionally, based on point-based annotations, the model's weakly supervised learning approach significantly reduces the need for detailed labeling, making it scalable for large-scale deployments with minimal annotation efforts. Experimental results demonstrate that AMSO-SFS outperforms existing SOTA models in terms of detection accuracy, real-time processing capabilities, and power efficiency, making it highly applicable for use on resource-limited drones. Its ability to operate in real-time on-edge devices such as the NVIDIA Jetson AGX Xavier ensures its practicality for continuous wildfire monitoring and early detection missions. However, while the model demonstrates robust performance across varied conditions, there remain areas for further refinement and adaptation.

Future work will focus on expanding AMSO-SFS's functionality and adaptability. One development area involves integrating additional environmental variables, such as wind speed, humidity, and temperature, to enable more accurate predictions of fire spread, which could provide valuable insights for firefighting strategies. There is also potential to adapt AMSO-SFS for other disaster detection scenarios, such as floods and landslides, by incorporating relevant sensor data and modifying the model's architecture. Additionally, enhancements in weakly supervised learning techniques, such as semi-supervised or active learning, could improve boundary precision, especially under dense smoke or fast-evolving fire conditions. Future research will explore advanced model compression methods, including pruning and quantization, as well as efficient transformer architectures, to support deployment on more affordable or smaller drones with limited processing capabilities. Expanding the range of sensors, such as incorporating LIDAR or hyperspectral sensors, could further bolster the model detection performance in extreme environments, allowing it to detect hotspots through heavy smoke or distinguish between smoke and other airborne particles. Another promising direction is adaptive real-time model updating to maintain detection accuracy across different regions and seasons. Techniques like edge-based continual learning could allow AMSO-SFS to adapt based on newly gathered data, ensuring high performance without extensive re-training. Finally, the development of a user-friendly interface for emergency response teams would enable effective in-field use of AMSO-SFS, with real-time visualizations, alert notifications, and fire spread forecasting, supporting rapid and informed decision-making.

AMSO-SFS represents a significant advancement in drone-based wildfire and smoke detection, combining high accuracy, real-time performance, and energy efficiency to support timely wildfire response efforts and minimize environmental damage. The outlined future work highlights the model's potential for further refinement and versatility, aiming to create a highly adaptive and scalable tool for comprehensive disaster management and environmental monitoring.

Author Contributions: Methodology, A.A., S.U., M.B.S., A.K., A.B., H.S.J. and M.M.; Software, A.A., S.U., M.B.S. and A.K.; Validation, S.U., A.A., M.B.S., A.K., A.B., H.S.J. and M.M.; Formal analysis, A.K., A.B., H.S.J., M.M. and S.U.; Resources, M.B.S., A.K., A.B. and M.M.; Data curation, H.S.J. and M.M.; Writing—original draft, A.A., S.U., M.B.S., A.K., A.B., H.S.J. and M.M.; Writing—review and editing, A.A., S.U., M.B.S., A.K., A.B., H.S.J. and M.M.; Supervision, H.S.J. and M.M.; Project administration,

M.B.S., A.K., A.B., H.S.J. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00412141).

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

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

References

1. Özel, B.; Alam, M.S.; Khan, M.U. Review of Modern Forest Fire Detection Techniques: Innovations in Image Processing and Deep Learning. *Information* **2024**, *15*, 538. [\[CrossRef\]](#)
2. Saleh, A.; Zulkifley, M.A.; Harun, H.H.; Gaudreault, F.; Davison, I.; Spraggon, M. Forest fire surveillance systems: A review of deep learning methods. *Heliyon* **2024**. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Muksimova, S.; Mardieva, S.; Cho, Y.I. Deep encoder–decoder network-based wildfire segmentation using drone images in real-time. *Remote Sens.* **2022**, *14*, 6302. [\[CrossRef\]](#)
4. Jonnalagadda, A.V.; Hashim, H.A. SegNet: A segmented deep learning based Convolutional Neural Network approach for drones wildfire detection. *Remote Sens. Appl. Soc. Environ.* **2024**, *34*, 101181. [\[CrossRef\]](#)
5. 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. [\[CrossRef\]](#)
6. Yandouzi, M.; Berrahal, M.; Grari, M.; Boukabous, M.; Moussaoui, O.; Azizi, M.; Ghomid, K.; Elmiad, A.K. Semantic segmentation and thermal imaging for forest fires detection and monitoring by drones. *Bull. Electr. Eng. Inform.* **2024**, *13*, 2784–2796. [\[CrossRef\]](#)
7. Ali, M.F.; Jayakody, D.N.K.; Muthuchidambaranathan, P. Revolutionizing Firefighting: UAV-Based Optical Communication Systems for Wildfires. *Photonics* **2024**, *11*, 656. [\[CrossRef\]](#)
8. Abdusalamov, A.; Umirzakova, S.; Safarov, F.; Mirzakhilov, S.; Egamberdiev, N.; Cho, Y.I. A Multi-Scale Approach to Early Fire Detection in Smart Homes. *Electronics* **2024**, *13*, 4354. [\[CrossRef\]](#)
9. Marques, T.; Carreira, S.; Miragaia, R.; Ramos, J.; Pereira, A. Applying deep learning to real-time UAV-based forest monitoring: Leveraging multi-sensor imagery for improved results. *Expert Syst. Appl.* **2024**, *245*, 123107. [\[CrossRef\]](#)
10. Redmon, J. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
11. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [\[CrossRef\]](#)
12. Yunusov, N.; Islam, B.M.S.; Abdusalomov, A.; Kim, W. Robust Forest Fire Detection Method for Surveillance Systems Based on You Only Look Once Version 8 and Transfer Learning Approaches. *Processes* **2024**, *12*, 1039. [\[CrossRef\]](#)
13. Sheng, S.; Liang, Z.; Xu, W.; Wang, Y.; Su, J. FireYOLO-Lite: Lightweight Forest Fire Detection Network with Wide-Field Multi-Scale Attention Mechanism. *Forests* **2024**, *15*, 1244. [\[CrossRef\]](#)
14. Sudiana, D.; Lestari, A.I.; Riyanto, I.; Rizkinia, M.; Arief, R.; Prabuwo, A.S.; Sri Sumantyo, J.T. A hybrid convolutional neural network and random forest for burned area identification with optical and synthetic aperture radar (SAR) data. *Remote Sens.* **2023**, *15*, 728. [\[CrossRef\]](#)
15. Shama, A.; Zhang, R.; Wang, T.; Liu, A.; Bao, X.; Lv, J.; Zhang, Y.; Liu, G. Forest fire progress monitoring using dual-polarisation Synthetic Aperture Radar (SAR) images combined with multi-scale segmentation and unsupervised classification. *Int. J. Wildland Fire* **2023**, *33*, WF23124. [\[CrossRef\]](#)
16. Mastro, P.; Masiello, G.; Serio, C.; Pepe, A. Change detection techniques with synthetic aperture radar images: Experiments with random forests and Sentinel-1 observations. *Remote Sens.* **2022**, *14*, 3323. [\[CrossRef\]](#)
17. Meng, L.; Yan, C.; Lv, S.; Sun, H.; Xue, S.; Li, Q.; Zhou, L.; Edwing, D.; Edwing, K.; Geng, X.; et al. Synthetic aperture radar for geosciences. *Rev. Geophys.* **2024**, *62*, e2023RG000821. [\[CrossRef\]](#)
18. Rokhmatuloh; Ardiansyah; Indratmoko, S.; Riyanto, I.; Margatama, L.; Arief, R. Burnt-Area Quick Mapping Method with Synthetic Aperture Radar Data. *Appl. Sci.* **2022**, *12*, 11922. [\[CrossRef\]](#)
19. Simões, J.V.O.; Negri, R.G.; Souza, F.N.; Mendes, T.S.G.; Bressane, A. Unsupervised burned areas detection using multitemporal synthetic aperture radar data. *J. Appl. Remote Sens.* **2024**, *18*, 014513. [\[CrossRef\]](#)
20. Amitrano, D.; Di Martino, G.; Guida, R.; Iervolino, P.; Iodice, A.; Papa, M.N.; Riccio, D.; Ruello, G. Earth environmental monitoring using multi-temporal synthetic aperture radar: A critical review of selected applications. *Remote Sens.* **2021**, *13*, 604. [\[CrossRef\]](#)
21. Mondini, A.C.; Guzzetti, F.; Chang, K.T.; Monserrat, O.; Martha, T.R.; Manconi, A. Landslide failures detection and mapping using Synthetic Aperture Radar: Past, present and future. *Earth-Sci. Rev.* **2021**, *216*, 103574. [\[CrossRef\]](#)

22. Dewald, A.; Mascarenhas, C.; Dykema, J.; Anderson, J. Feasibility and Sensitivity Analysis of Integrated Synthetic Aperture Radar (SAR) on HALE Aircraft for Wildfire Applications. Aerospace Research Center. In Proceedings of the AIAA Aviation Forum and Ascend, Las Vegas, NV, USA, 29 July–2 August 2024; p. 4298.
23. Rahman, H.I.; Saad, A.F.; Yani, A. Drone Based Fire Detection System Based on Convolutional Neural Network. *Int. J. Artif. Intell.* **2024**, *11*, 26–36. [[CrossRef](#)]
24. Li, Y.; Lv, X.; Huang, P.; Xu, W.; Tan, W.; Dong, Y. SAR Ship target detection based on improved YOLOv5s. In Proceedings of the 2021 International Conference on Control, Automation and Information Sciences (ICCAIS), Xi'an, China, 14–17 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 354–358.
25. Cheng, G.; Yang, J.; Gao, D.; Guo, L.; Han, J. High-quality proposals for weakly supervised object detection. *IEEE Trans. Image Process.* **2020**, *29*, 5794–5804. [[CrossRef](#)] [[PubMed](#)]
26. Pan, J.; Ou, X.; Xu, L. A collaborative region detection and grading framework for forest fire smoke using weakly supervised fine segmentation and lightweight faster-RCNN. *Forests* **2021**, *12*, 768. [[CrossRef](#)]
27. Makhmudov, F.; Umirzakova, S.; Kutlimuratov, A.; Abdusalomov, A.; Cho, Y.I. Advanced Object Detection for Maritime Fire Safety. *Fire* **2024**, *7*, 430. [[CrossRef](#)]
28. Vaidya, H.; Gupta, A.; Ghanshala, K.K. Attention-driven YOLOv5 for wildfire smoke detection from camera images. *J. Earth Syst. Sci.* **2024**, *133*, 174. [[CrossRef](#)]
29. Luo, J.; Yang, X.; Yu, Y.; Li, Q.; Yan, J.; Li, Y. PointOBB: Learning Oriented Object Detection via Single Point Supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 16730–16740.
30. Hopkins, B.; O'Neill, L.; Afghah, F.; Razi, A.; Rowell, E.; Watts, A.; Fule, P.; Coen, J. FLAME 2: Fire Detection and Modeling: Aerial Multi-Spectral Image Dataset. 2022. Available online: <https://ieee-dataport.org/open-access/flare-2-fire-detection-and-modeling-aerial-multi-spectral-image-dataset> (accessed on 20 December 2023).
31. Shamsoshoara, A.; Afghah, F.; Razi, A.; Zheng, L.; Fulé, P.; Blasch, E. The FLAME Dataset: Aerial Imagery Pile Burn Detection Using Drones (UAVs). 2020. Available online: <https://ieee-dataport.org/open-access/flare-dataset-aerial-imagery-pile-burn-detection-using-drones-uavs> (accessed on 20 December 2023).

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.