fire

Article

Revolutionizing Wildfire Detection Through UAV-Driven Fire Monitoring with a Transformer-Based Approach

Shakhnoza Muksimova [1], Sabina Umirzakova [1,*], Sevara Mardieva [1], Mirjamol Abdullaev [2] and Young Im Cho [1,*]

[1] Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 461-701, Gyeonggi-do, Republic of Korea; shakhnoza02@gachon.ac.kr (S.M.); sevara1998@gachon.ac.kr (S.M.)
[2] Department of Information Systems and Technologies, Tashkent State University of Economics, Tashkent 100066, Uzbekistan; abdullaev@tsue.uz
* Correspondence: sabinatuit@gachon.ac.kr (S.U.); yicho@gachon.ac.kr (Y.I.C.)

Abstract: The rapid detection and accurate localization of wildfires are critical for effective disaster management and response. This study proposes an innovative Unmanned aerial vehicles (UAVs)-based fire detection system leveraging a modified Miti-DETR model tailored to meet the computational constraints of drones. The enhanced architecture incorporates a redesigned AlexNet backbone with residual depthwise separable convolution blocks, significantly reducing computational load while improving feature extraction and accuracy. Furthermore, a novel residual self-attention mechanism addresses convergence issues in transformer networks, ensuring robust feature representation for complex aerial imagery. The model, which was trained on the FLAME dataset encompassing diverse fire scenarios, demonstrates superior performance in terms of Mean Average Precision (mAP) and Intersection over Union (IoU) metrics compared to existing systems. Its capability to detect and localize fires across varied backgrounds highlights its practical application in real-world scenarios. This advancement represents a pivotal step forward in applying deep learning for real-time wildfire detection, with implications for broader emergency management applications.

Keywords: wildfire detection; object detection; drone; deep learning; transformers in image processing; wildfire management; real-time imaging; environmental monitoring

check for updates

Citation: Muksimova, S.; Umirzakova, S.; Mardieva, S.; Abdullaev, M.; Cho, Y.I. Revolutionizing Wildfire Detection Through UAV-Driven Fire Monitoring with a Transformer-Based Approach. *Fire* **2024**, *7*, 443. https://doi.org/10.3390/fire7120443

Academic Editor: Robert Kremens

Received: 21 October 2024
Revised: 22 November 2024
Accepted: 27 November 2024
Published: 28 November 2024

1. Introduction

Wildfires have emerged as one of the most devastating natural disasters, causing extensive damage to ecosystems, property, and human life. Their increasing frequency and intensity, exacerbated by climate change, demand innovative approaches to detection and management [1]. According to the Korea Forest Service, 596 wildfire incidents occurred in South Korea in 2023 alone, affecting over 5000 hectares [2]. This underscores the urgent need for advanced, real-time monitoring systems to mitigate such disasters effectively. Traditional methods of wildfire detection, including ground-based surveillance and satellite imaging, often fall short of providing the rapid and accurate detection required for timely interventions, as shown in Figure 1.

Drones, equipped with advanced sensors and imaging technologies, have emerged as a transformative solution, offering a unique aerial perspective that significantly enhances fire detection capabilities [3]. Domestically, research in South Korea has focused on integrating UAVs with traditional monitoring systems to improve fire detection in forested regions. For instance, studies have explored combining drones with IoT networks to enhance data collection and transmission [4]. While these approaches have demonstrated potential, their reliance on conventional imaging and algorithmic frameworks limits their scalability and precision in real-time scenarios. Ref. [5] introduces a novel fire and smoke detection method utilizing capsule networks with attention feature maps, designed to improve the accuracy of incident classification in outdoor video recordings. This approach demonstrates high performance in real-time applications and robustness under diverse conditions. Recent advancements in UAVs technology have demonstrated their potential to surpass traditional methods of wildfire detection, such as ground-based surveillance and satellite imaging. This paper [6] introduces a deep encoder-decoder network for real-time wildfire segmentation using drone imagery. This network employs attention mechanisms and residual blocks to enhance both accuracy and efficiency, thereby outperforming existing models in the detection and segmentation of wildfires and smoke. UAVs can navigate challenging terrain, providing dynamic, high-frequency data that significantly enhances situational awareness during wildfire events. Studies have shown that drones equipped with RGB cameras, thermal imaging systems, and LiDAR sensors can detect fires at various stages, from smoldering embers to active flames. Ref. [7] conducted a study on thermal imaging, highlighting its proven effectiveness in detecting heat signatures under low-visibility conditions, such as those encountered in dense smoke or nighttime operations. LiDAR technology complements this by generating precise 3D spatial data, which is crucial for mapping terrain and predicting fire spread dynamics. In addition to sensor advancements, integrating AI algorithms into UAVs-based systems has revolutionized wildfire detection and prediction capabilities. Machine learning models, including convolutional neural networks (CNNs) and transformer-based architectures, enable UAVs to process vast amounts of data in real time, improving detection accuracy and operational efficiency. Recent studies, such as those leveraging YOLOv5 and Detectron2, have demonstrated a strong performance in object detection tasks, though they often face challenges in handling complex backgrounds or resource constraints when deployed on UAVs [8,9]. Incorporating AI algorithms capable of analyzing multi-modal data, such as RGB, thermal, and LiDAR inputs, has the potential to further refine fire identification systems.

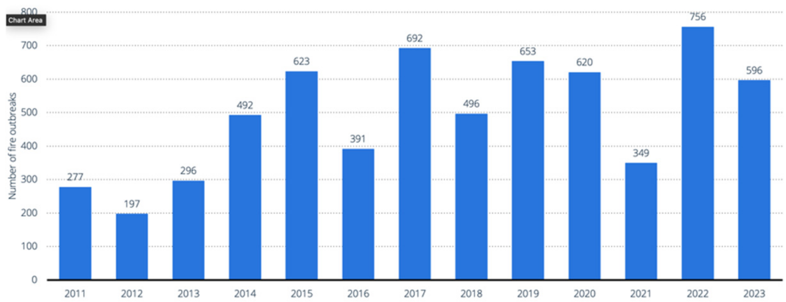

Figure 1. Forest fire incidents in South Korea from 2011 to 2023 [2].

Despite these advancements, significant challenges remain, particularly in balancing computational efficiency with detection accuracy. UAVs are inherently resource-constrained, requiring lightweight, energy-efficient models to operate effectively. This study addresses these challenges by introducing a novel fire detection framework based on the modified Miti-DETR model. By incorporating a redesigned AlexNet backbone with residual depthwise separable convolution blocks, the proposed model significantly reduces computational load while enhancing feature extraction. A novel residual self-attention mechanism is also integrated to address convergence issues commonly associated with deep transformer networks, ensuring robust feature representation.

This capability is crucial, particularly in remote or inaccessible regions where traditional monitoring methods are either impractical or inefficient. However, drone potential extends beyond mere observation; they can be pivotal in implementing sophisticated fire detection models that leverage artificial intelligence (AI) to predict and prevent fire outbreaks [10]. Given the constraints of drone hardware capabilities and the urgency of fire detection tasks, the core challenge in deploying such models lies in their need to be both highly accurate and extremely efficient [11].

This study addresses these challenges by introducing a novel fire detection model tailored for UAVs deployment. The proposed model, a modified Miti-DETR, combines the strengths of CNNs and transformer-based architectures to achieve high precision and efficiency [12]. The innovation lies in the redesigned AlexNet backbone, which incorporates residual depthwise separable convolution blocks to reduce computational load while enhancing feature extraction. Additionally, a novel residual self-attention mechanism ensures robust feature representation, mitigating convergence issues commonly associated with deep transformer networks. By integrating these modifications, the model achieves superior performance in detecting and localizing fire elements across varied and complex aerial imagery.

The contributions of this study are threefold. First, it introduces a lightweight, high-performance fire detection model specifically designed to address the hardware constraints of UAVs. Second, it demonstrates the model's capability to outperform existing systems in terms of accuracy and convergence speed, validated through a comprehensive evaluation of the FLAME dataset. Third, it establishes the broader applicability of the model in real-world wildfire management scenarios, highlighting its potential for improving disaster response strategies.

2. Literature Review

2.1. Methods of Visual Detection

The value of vision-based techniques, recognized for their capabilities in real-time data capture, broad detection scope, and straightforward verification and recording, has established them as central in the field of wildfire monitoring and detection [13]. Recently, several researchers have explored the properties of remote-sensing images, introducing methods for high-resolution analysis [14,15]. A detailed review in [16] examined object detection methodologies, particularly the advancements and applications of the YOLO algorithm along with relevant datasets. Ref. [17] presents an advanced fire and smoke detection model that integrates the Vision Transformer (ViT) with YOLOv5s, enhancing detection in environments with obscured or widespread fires by effectively processing both local and global image features. In [18], an innovative approach tailored for smart city environments utilized YOLOv8 to improve fire detection accuracy. The research in [19] leveraged the Detectron2 model alongside deep learning techniques to increase the reliability of wildfire detection. In [20], efforts focused on enhancing the detection of wildfires and smoke, emphasizing their importance in wildfire management. Additionally, ref. [21] discussed the use of UAVs to enhance wildfire surveillance, focusing on precise fire identification and localization.

Given the intricate and unpredictable nature of wildfires, integrating multiple data sources from various locations is crucial. When managing extensive or multiple wildfires, the update rate from a single drone may prove inadequate. Common attributes analyzed in existing studies include the geometry, color, and motion of fires. Specifically, color features, often extracted through trained networks, are used to delineate fire zones. The work in [22] presented a multi-UAV system that effectively detects wildfires, detailing the components of helicopter UAVs and employing motion and color analysis to obtain clear images. Fire characteristics, extending beyond color to irregular shapes and consistent motion at specific spots, were considered. This study also introduced algorithms tailored for ultraviolet and visual spectrum cameras. In [23], a novel methodology combining color-motion-shape features with machine learning was introduced to detect fires. A robust wildfire detection system was proposed in [24], necessitating the precise classification of fire imagery against non-fire images. This involved curating a diverse dataset (Deep-Fire) and employing VGG-19 transfer learning to enhance prediction accuracy, while comparing various machine learning techniques. Additionally, ref. [25] introduced a deep learning-based model for wildfire detection from satellite imagery, using RCNNs to decrease prediction times and achieve a 97.29% accuracy rate in identifying fires in remote areas.

2.2. UAVs for Monitoring and Detecting

Aerial surveillance of fires, while effective, is costly and involves considerable risks, particularly with uncontrolled wildfires. As noted in [26], the advancement of drone technology in firefighting primarily focuses on remote sensing capabilities for forest fires. The study by [27] introduced an autonomous system for forest fire monitoring designed to swiftly locate and track critical hot spots. This system incorporates an algorithm that directs drones to these hot spots, demonstrated through a mixed reality experiment that simulated forest fire progression and compared the algorithm's effectiveness against basic fire tracking strategies. The development of a real-time wildfire monitoring system is further exemplified in [28], where UAVs, equipped with various sensors, mini processors, and cameras, process data from onboard sensors and imagery to provide timely information on fire dynamics.

Beyond mere remote sensing, drones have potential applications in direct firefighting tasks, such as igniting controlled burns from the air, although the practical deployment of such technologies is still emerging due to logistical challenges related to the transport of significant volumes of water and fire retardants [29]. In [30], a novel framework combining the YOLOv4 deep network with LiDAR technology was outlined, presenting a cost-effective method for managing UAV-based fire detection systems.

2.3. Post-Fire Recovery and Long-Term Ecological Monitoring

Beyond immediate detection, UAVs have shown great promise in post-fire recovery and ecological monitoring. The severity and frequency of wildfires have long-lasting impacts on ecosystems, affecting vegetation regrowth, soil health, and biodiversity. Studies by [31] revealed that fire severity directly influences the rate of ecosystem recovery, with low-severity fires often resulting in quicker regrowth compared to high-severity fires that cause extensive soil degradation. UAVs equipped with multi-spectral and thermal sensors play a pivotal role in monitoring these recovery processes. They enable researchers to map burned areas, assess soil erosion, and track vegetation regrowth over time. Research by [32] has demonstrated the utility of UAVs in post-fire recovery, particularly in analyzing changes in vegetation health and soil composition. These tools provide valuable insights into the long-term ecological effects of wildfires, helping inform restoration and conservation strategies. Wildfires not only pose ecological challenges but also have significant societal implications. The economic costs of wildfires, including damage to property and infrastructure, highlight the need for advanced monitoring systems that can mitigate these impacts. Ref. [33] argues that integrating UAVs with AI-driven models and advanced sensors is essential for improving both the efficiency and accessibility of wildfire management practices. Additionally, the societal benefits of early detection systems, such as reducing evacuation times and protecting vulnerable communities, underscore the importance of continued innovation in this field.

This paper details the development and implementation of the Miti-DETR model for wildfire detection using drones. We explore the model architecture, modifications, and the rationale behind these changes, emphasizing the balance between computational efficiency and detection performance. By deploying this model in drones equipped with advanced imaging technologies, we aim to significantly reduce the response time in wildfire management and enhance the accuracy of fire detection, ultimately contributing to better resource allocation and more effective mitigation strategies.

3. Methodology

We have modified the widely recognized detection model, Miti-DETR, which has demonstrated remarkable performance as shown in Figure 2. To enhance its efficiency, we replaced the backbone of the model with AlexNet and incorporated the residual depthwise separable convolution block in the modified AlexNet architecture, as illustrated in Figure 3.

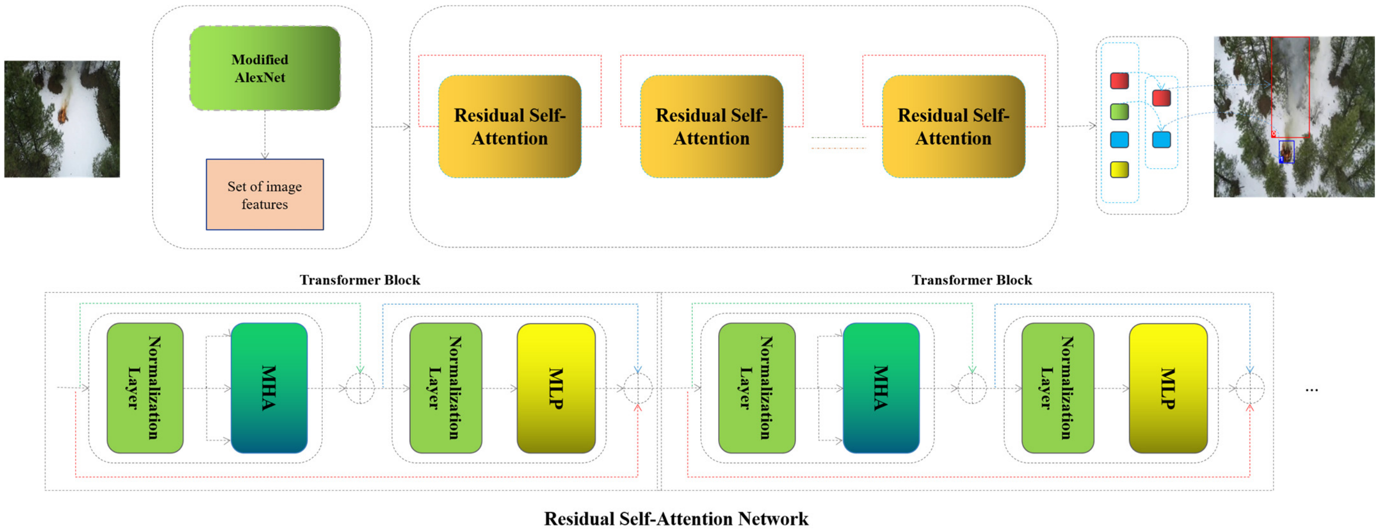

Figure 2. Modified Miti-DERT framework.

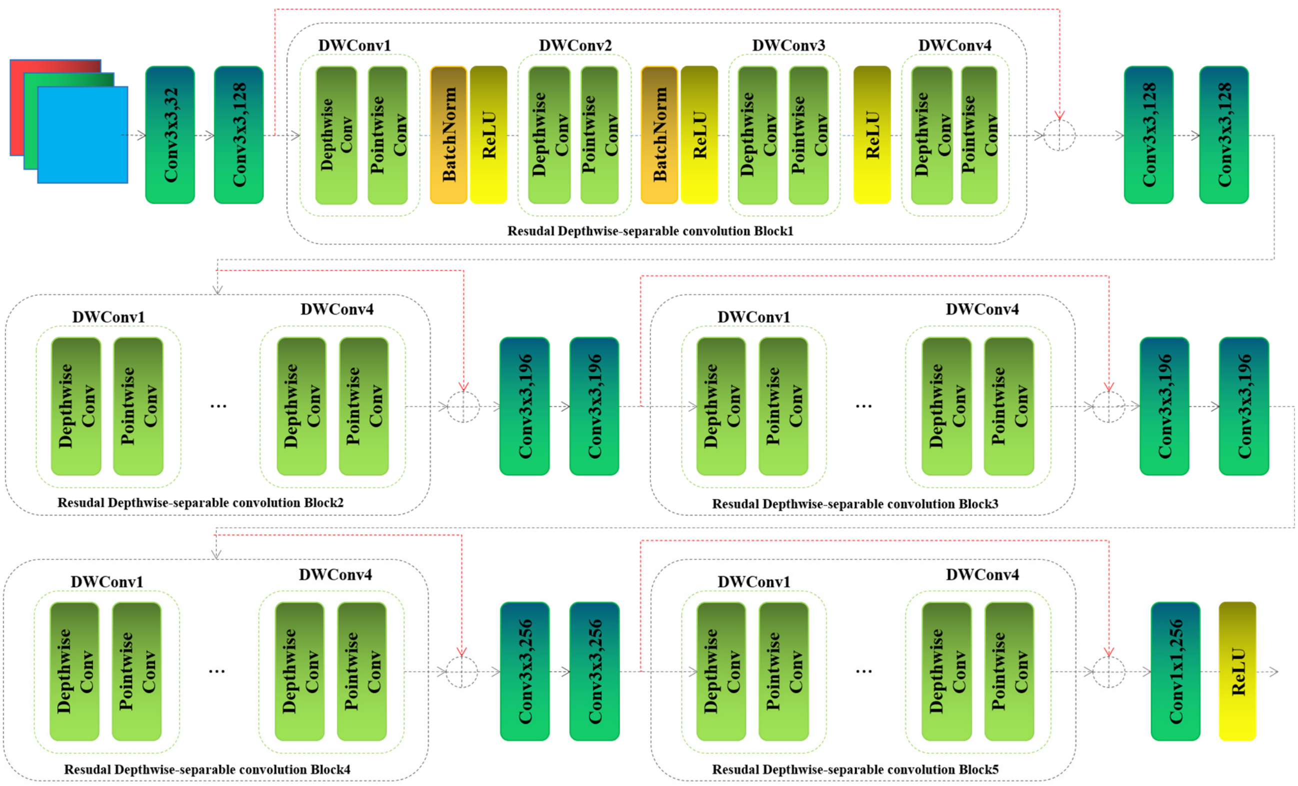

Figure 3. Modified AlexNet.

3.1. Miti-DETR

Miti-DETR [34] represents an innovative approach to object detection, building upon the fundamental architecture of the original DETR model. The architecture integrates the CNNs as a feature extractor and utilizes transformer-based encoder-decoder networks for object detection. However, Miti-DETR distinguishes itself by introducing a residual self-attention mechanism that addresses limitations observed in standard transformers, particularly the tendency of pure attention networks to degenerate and lose expressive power as they deepen, as shown in Figure 2. This residual self-attention network allows for direct connections between the inputs and outputs of each transformer layer, ensuring that non-attention-based information participates in the attention propagation process.

This structural innovation mitigates the convergence issues common in attention networks, slowing down their rate of convergence and thereby preventing the feature collapse seen in deeper networks. By diversifying the path distribution of parameter updates, Miti-DETR not only stabilizes attention mechanisms but also enhances the learning process. Moreover, the residual connections of the model maximize the diversity of attention features, helping counteract the natural inductive bias toward token uniformity in self-attention networks. This leads to better feature representation and stronger performance in object detection tasks. Miti-DETR has demonstrated significant improvements in both detection accuracy and convergence speed when evaluated on the challenging FLAME dataset. Additionally, its design allows for easy generalization to other models without requiring extensive modifications, making it a versatile tool in transformer-based object detection.

3.2. AlexNet

AlexNet, a pioneering deep learning architecture, was introduced by Alex Krizhevsky [35] in 2012 and played a crucial role in establishing the dominance of CNNs in the field of computer vision. The architecture stood out due to its ability to achieve state-of-the-art performance on the ImageNet dataset, an extensive image classification challenge, by reducing the top-5 error rate from 26% to 15.3%. The design of the model, though simple by standards, is revolutionary in both scale and depth, marking a significant departure from traditional machine learning techniques. The network consists of eight layers, five of which are convolutional layers and three are fully connected layers. The convolutional layers are responsible for detecting patterns and features in images, such as edges, textures, and object parts. These layers are followed by max-pooling layers, which help reduce the spatial dimensions of the feature maps while preserving the most essential information. This approach allows the network to capture hierarchical representations of objects, progressively learning from simple features to complex ones. One of the key innovations of AlexNet was its use of ReLU [36] as the activation function, which sped up training significantly compared to traditional sigmoid or tanh activations. ReLU helped address the vanishing gradient problem and allowed deeper networks to train more efficiently by preventing gradients from becoming too small. Moreover, AlexNet has made effective use of GPU acceleration, which was a relatively novel idea at the time. Training such a large network on CPUs would have been prohibitively slow, but by leveraging the parallel processing capabilities of GPUs, the training time is drastically reduced, enabling the network to be scaled up without sacrificing practicality.

3.3. The Proposed Model

In our work, we modified the backbone of the Miti-DETR by replacing the CNNs with a modified AlexNet. To make AlexNet a more powerful feature extractor, we made it deeper by adding five residual depthwise-separable convolution blocks, which consist of batch normalizations and ReLU activation functions, as shown in Figure 3. In the initial baseline model, there were seven convolutional layers and two flattened layers. However, since our task was object detection, we removed the flattened layers and replaced them with a single final convolutional layer. The input image $X_{i\_img} \in R^{W\times H\times C}$ is first processed to extract low-level features using two convolutional layers.

$$F_{conv\_layer1} = F_{3\times 3}(F_{3\times 3}(X_{i\_img})) \tag{1}$$

where, F_{conv_layer1} represents the first two convolution layers for extracting the low-level features, with kernel size 3×3.

$$\begin{gathered} F_{res\_block1} = F_4(F_3(F_2(F_1(F_{conv\_layer1})))), \\ F_1 = max(0,\ BatchNorm(F_{DW1}(F_{conv\_layer1}))), \\ F_2 = max(0,\ BatchNorm(F_{DW2}(F_1))), \\ F_3 = max(0,\ (F_{DW3}(F_2))), \end{gathered} \tag{2}$$

$$F_4 = F_{DW4}(F_3)$$

F_{res_block1} represents the full structure of the residual depthwise-separable convolution block, comprising four internal depthwise and pointwise convolution layers, F_1, F_2, F_3, and F_4. Moreover, batch normalization and ReLU activation functions are incorporated. Batch normalization is used to normalize the input to each layer, improving training stability and accelerating convergence by mitigating internal covariate shifts. ReLU is employed as an activation function to introduce non-linearity, prevent vanishing gradients, and ensure efficient learning in deep networks.

$$F_{concat} = Concat(F_{res\_block1} + F_{conv\_layer1}) \tag{3}$$

$$F_{conv\_layer2} = F_{3\times3}(F_{3\times3}(F_{concat})) \tag{4}$$

Following the first residual block, the input and output of the block are concatenated to prevent information loss and preserve important features. After this, a second set of two convolutional layers is applied to further extract features while increasing the number of channels, enhancing the capacity of the model to capture more complex patterns and representations.

$$F_{res\_block2} = Concat(F_4(F_3(F_2(F_1(F_{conv\_layer2})))),\ F_{conv\_layer2}) \tag{5}$$

$$F_{conv\_layer3} = F_{3\times3}(F_{3\times3}(F_{res\_block2})) \tag{6}$$

The same process is repeated for the subsequent 1 to N residual blocks, where N = 5. Each residual block is followed by double convolution layers, maintaining the pattern of feature extraction and channel expansion as the model deepens. This consistent structure allows for more complex hierarchical representations to be captured throughout the network.

$$\begin{gathered}
F_{res_block3} = Concat(F_4(F_3(F_2(F_1(F_{conv_layer3})))),\ F_{conv_layer3}) \\
F_{conv_layer4} = F_{3\times3}(F_{3\times3}(F_{res_block3})) \\
F_{res_block4} = Concat(F_4(F_3(F_2(F_1(F_{conv_layer4})))),\ F_{conv_layer4}) \\
F_{conv_layer5} = F_{3\times3}(F_{3\times3}(F_{res_block4})) \\
F_{res_block5} = Concat(F_4(F_3(F_2(F_1(F_{conv_{layer5}})))),\ F_{conv_{layer5}}) \\
F_{conv_layer6} = max(0,\ F_{1\times1}(F_{res_block5}))
\end{gathered} \tag{7}$$

All the captured coarse features, resulting from the application of the ReLU activation function, are then passed to the MITI-DETR model. ReLU ensures that non-linearities are introduced while preserving critical features, allowing the MITI-DETR to effectively process these features for further detection and recognition tasks.

3.4. The Metrics and Losses

mAP is a prevalent metric utilized in object detection for evaluating model precision. It calculates the average precision across all object categories and various intersection-over-union (IoU) thresholds, yielding a thorough assessment of model accuracy. For each category, precision and recall are determined at different confidence levels, with Average Precision (AP) defined as the area under the precision-recall curve. Subsequently, mAP is derived as the average of the AP values for all classes, offering a consolidated overview of model performance. In this context, N denotes the total number of object classes, while $AP(i)$ indicates the Average Precision computed for the specific class i.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \tag{8}$$

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (9)$$

IoU measures the extent of overlap between the predicted bounding box and the ground truth bounding box. The values of IoU range from 0 to 1, with higher values indicating a closer alignment between the predicted and actual object. A commonly used threshold for IoU is 0.5, representing 50% overlap, to classify a prediction as correct. In this context, $|A \cap B|$ denotes the area of overlap between the predicted and ground truth bounding boxes, while $|A \cup B|$ signifies the area of the union of these two boxes.

$$precision = \frac{TP}{TP + FP} \quad (10)$$

$$recall = \frac{TP}{TP + FN} \quad (11)$$

Precision is the proportion of true positive detections relative to the total number of positive predictions, which includes both true positives and false positives. Recall is the proportion of true positive detections relative to the total number of actual ground truth objects, encompassing both true positives and false negatives.

$$L_{CE} = -\sum_{i=1}^{C} y_i log(y\prime_i) \quad (12)$$

$$\int_{match} = \lambda_{class} L_{CE} + \lambda_{bbox} L_{L1} + \lambda_{giou} L_{GIoU} \quad (13)$$

Classification Loss (Cross-Entropy Loss) is a loss function which is employed to ensure accurate object classification. It imposes penalties on incorrect predictions, driving the model to reliably determine the correct class for detected objects. C is the total number of classes, y_i is the ground truth probability, and $y\prime_i$ is the predicted probability for the *i-th* class. Hungarian Matching Loss is a crucial component of DETR-based models, including MITI-DETR. The loss function utilizes the Hungarian algorithm to perform bipartite matching between predicted and ground truth boxes, ensuring an optimal one-to-one correspondence between predictions and actual objects. L_{CE} is the cross-entropy classification loss, L_{L1} is the L1 loss for bounding box regression, L_{GIoU} is the GIoU loss for bounding boxes, and λ_{class}, λ_{giou}, and λ_{bbox} are hyperparameters that balance the contribution of each component.

4. Experimental Setup and Results

4.1. The Dataset

The FLAME dataset [37] is a comprehensive and specialized resource designed specifically for fire detection tasks, supporting the development of accurate and reliable models. It includes a wide variety of images and videos that capture fire occurrences across both natural and urban environments, such as forest fires, industrial fires, and urban incidents. The dataset is diverse in terms of conditions, including different times of day, weather types, and varying fire intensities, which makes it suitable for training models that can handle real-world fire detection challenges. The FLAME dataset is meticulously annotated with detailed labels for fire, smoke, and non-fire regions, enabling fine-grained object detection. This precision allows models to effectively distinguish between fire-related phenomena and the background. The high resolution and quality of the images and videos in the dataset ensure that even subtle or small-scale fires can be detected. The data is captured from a variety of perspectives, including aerial views from drones and ground-level shots, enhancing its applicability to diverse fire monitoring systems, as shown in Figure 4.

Figure 4. Description of FLAME dataset examples: (**a**) non-fire and (**b**) fire.

The inclusion of both fire and non-fire images ensures that models can be trained without bias, improving their ability to differentiate between fire events and other visually similar situations. Moreover, the dataset is highly suitable for real-time fire detection model development, as it includes video sequences that allow for temporal analysis. This makes it ideal for applications that require rapid detection and response. With its rich variety and detailed annotations, the dataset offers an excellent benchmark for evaluating fire detection systems and helps ensure that models trained on it are robust and effective in a wide range of real-world scenarios.

4.2. Training Detail

The model is trained on a Tesla V100 GPU with a batch size of 16 for optimal computational efficiency, completing 50 epochs with an initial learning rate of 1×10^{-4}. This rate is periodically reduced by a factor of 0.1 every 15 epochs through a step decay schedule to enhance training effectiveness.

The Adam optimizer [38], with a weight decay of 1×10^{-5}, is used to minimize overfitting. The model employs a combination of cross-entropy for classification, L1 loss for bounding box regression, and GIoU loss to improve precision in object detection, which is crucial for the demanding accuracy needed in fire detection.

4.3. Data Pre-Processing

The preprocessing of the dataset begins with resizing the aerial images to ensure consistent input dimensions across all samples. Since the images are captured from varying perspectives and angles, standardizing their size is essential to ensure compatibility with deep learning models. Each image undergoes normalization, with pixel values scaled to a standardized range between zero and one. This step is crucial to stabilize the training process of the model by eliminating biases introduced by the varying intensity of grayscale values. With the dynamic nature of the scenes, particularly involving potential fire or smoke in forested regions, contrast adjustments are applied to enhance the visibility of critical features. Histogram equalization is employed to bring out the details in regions that may have low contrast, such as smoke against a cloudy sky or subtle variations in fire intensity. This ensures that even the faintest signals in the image are highlighted for the model to learn effectively. To address the potential variability in the angles from which the images are captured, data augmentation is incorporated as an integral step. Techniques such as rotations, flips, and slight translations are applied to the dataset to increase its diversity, ensuring that the model can generalize to various orientations and perspectives. This augmentation allows the model to handle real-world scenarios where fire or smoke might be captured from unpredictable angles, as shown in Figure 5. Noise reduction is another important step, particularly because shadows from trees or variations in lighting conditions can introduce artifacts into the images. A series of filters are applied to reduce this noise without losing important details, ensuring that the model focuses on the most relevant features for fire detection. Finally, each image is carefully annotated, either with bounding boxes or segmentation masks, to mark regions of interest, such as fire or smoke areas.

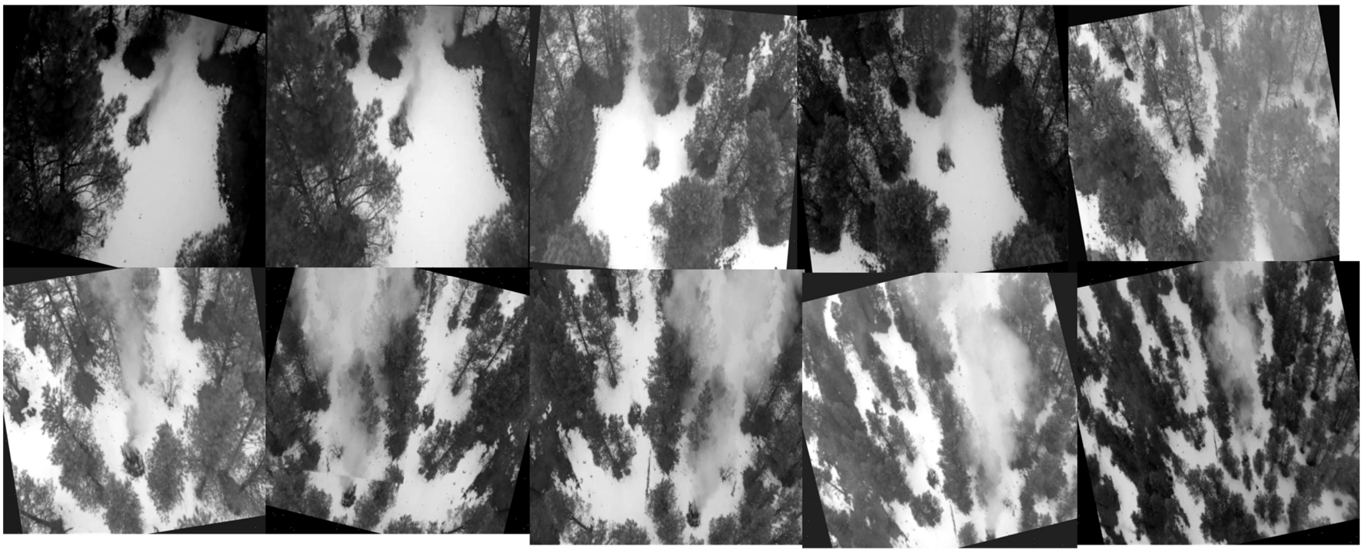

Figure 5. Examples of data augmentation.

These labels are crucial for guiding the learning process of the proposed model, enabling accurate detection and localization of fire-related elements in complex aerial environments.

4.4. Experiment and Results

Figure 6 shows the effectiveness of the modified MITI-DETR model for fire detection, particularly in aerial imagery captured by drones. The red bounding boxes clearly highlight regions where the model has detected smoke within the dense forest environment. These detections are precise and well-aligned with the ground truth annotations, showcasing the capability of the model to accurately identify areas of interest, despite the complex background of trees, snow, and varying lighting conditions. Moreover, the presence of multiple blue annotations within the images suggests that the model is also performing localization and classification tasks, possibly distinguishing between fire and smoke. The distribution of the bounding boxes across various regions demonstrates the robustness of the model in detecting fire across different perspectives and scales, which is crucial in aerial fire detection where the distance between the camera and the target may vary significantly. Figure 5 also illustrates the ability to handle occlusions and subtle visual cues, such as smoke partially obscured by tree canopies. The detected fire regions are not limited to easily visible flames, but also include subtle smoke plumes, which indicates the fine-tuned sensitivity of the proposed model to early fire indicators, an essential trait for real-time fire detection in drone surveillance applications. In the rest of the paper, these results reflect the strong performance of the proposed model in accurately detecting and localizing fire in challenging, real-world environments, effectively addressing the complexities involved in aerial fire detection tasks. This demonstrates the successful application of the modified MITI-DETR model in capturing critical fire-related information from drone-captured data, providing reliable outputs that can support rapid intervention and response efforts in wildfire management.

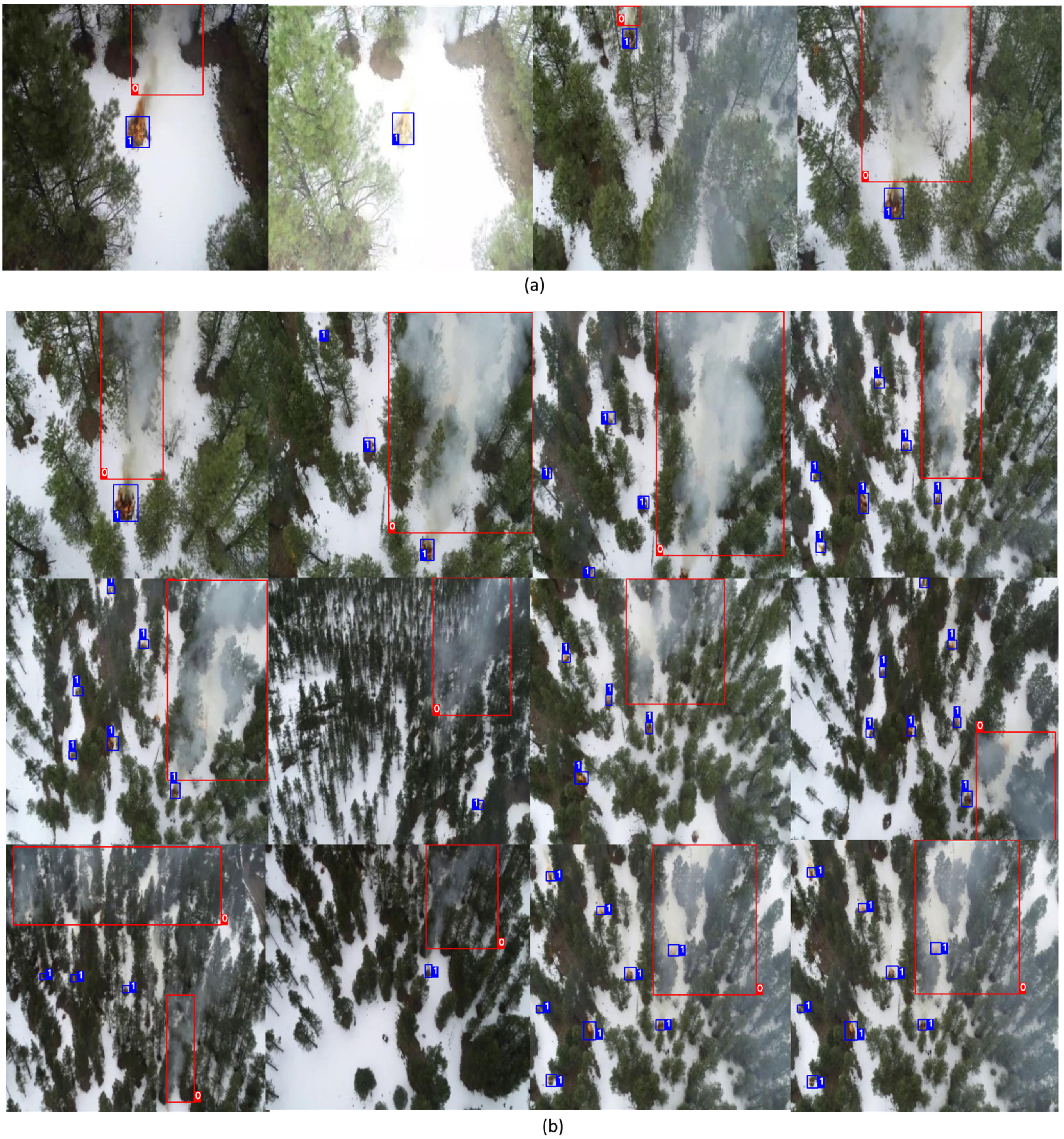

Figure 6. Examples of proposed model results: (**a**) Detection of smoke in snowy forest environments. (**b**) Identification of fire regions partially obscured by tree canopies.

4.5. Comparative Performance Analysis of Object Detection Models

Table 1 and Figure 7 presents a comparative evaluation of four models, such as the baseline, YOLOv5s, YOLOv5m, and the proposed model; all have been trained for 300 epochs. The performance metrics, including *AP* at different thresholds and object scales, are used to assess the efficacy of these models in object detection tasks. The overall *AP* of the baseline model stands at 36.3, while YOLOv5s scores slightly lower at 35.8. YOLOv5m performs similarly to the baseline with an AP of 36.2, but the proposed model surpasses all others, achieving an AP of 37.7. This indicates the proposed model demonstrated enhanced performance in detecting objects across different categories. At the IoU threshold of 50% AP_{50}, the baseline model achieves a score of 53.89, while YOLOv5s slightly underperforms with 53.12, as shown in Figure 6. YOLOv5m improves upon this with 54.47, but the proposed model outshines all others with an AP_{50} of 54.78, demonstrating its superior accuracy in detecting objects with more overlap between predicted and ground truth boxes. For the stricter 75% IoU threshold AP_{75}, the baseline model records 34.67, YOLOv5s drops to 32.01, and YOLOv5m scores 33.89. The proposed model, once again, leads with 34.78, reflecting its better precision when bounding box predictions need to closely match ground truth. For the detection of small objects AP_S, the baseline model achieves an AP of 15.5, while YOLOv5s underperforms slightly at 13.6. YOLOv5m improves marginally with 14.43, but the proposed model excels with an AP of 15.87, showing its superior capability in identifying small-scale objects in the dataset.

Table 1. Comparative Performance of Object Detection Models.

Model	Epochs	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	Precision	Recall	Inference Time (ms)	FLOPs
Baseline	300	36.3	53.89	34.67	15.5	33.11	53.7	84.5	82.1	10.2	32.8
YOLOv5s	300	35.8	53.12	32.01	13.6	31.35	52.0	83.7	80.9	12.3	28.5
YOLOv5m	300	36.2	54.47	33.89	14.43	31.78	53.98	85.2	83.0	13.5	35.2
Proposed Model	300	37.7	54.78	34.78	15.87	34.0	55.14	86.4	84.6	9.8	27.3

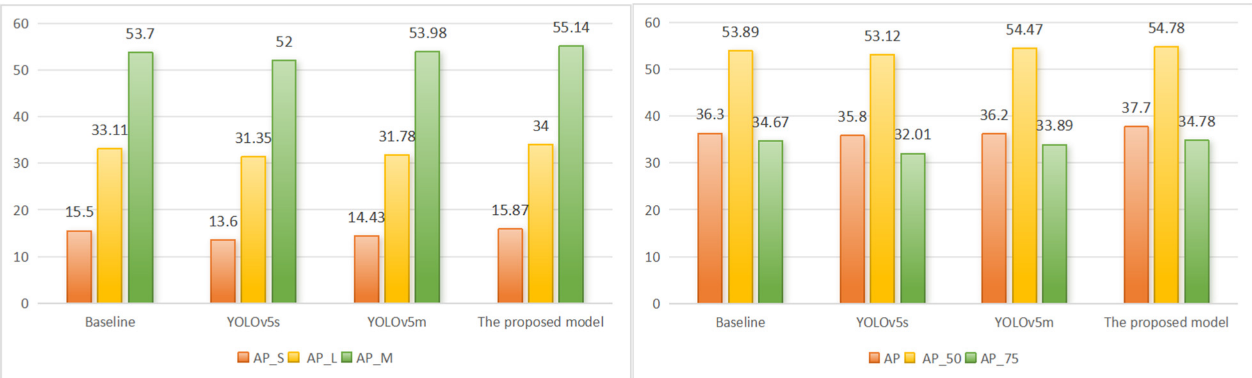

Figure 7. Evaluating YOLO Models and a Proposed Enhancement: Precision Metrics Across Object Scales.

For medium-sized objects AP_M, the baseline scores 33.11, YOLOv5s records a lower score of 31.35, while YOLOv5m performs slightly better at 31.78. The proposed model surpasses all others with a score of 34.0, highlighting its enhanced precision for medium-sized object detection. When it comes to large objects AP_L, the baseline model achieves a score of 53.7, YOLOv5s falls short at 52.0, while YOLOv5m improves slightly to 53.98. The proposed model outperforms the rest with an AP_L of 55.14, demonstrating its robustness in detecting larger objects. Moreover, the proposed model consistently demonstrates superior performance across most metrics, excelling in both overall precision and at various IoU thresholds, as well as across different object scales. These results underscore the effectiveness of the proposed modifications, particularly in tasks like fire detection using drone-captured imagery.

5. Discussion

This study introduces a significant advancement in UAV-based wildfire detection through the modified Miti-DETR model, which emphasizes computational efficiency and detection accuracy. By integrating a redesigned AlexNet backbone with residual depthwise separable convolution blocks, the model effectively addresses the hardware constraints of UAVs while enhancing precision in detecting and localizing fire elements. The results demonstrate that the proposed model outperforms traditional detection systems, such as YOLOv5s and YOLOv5m, across multiple metrics, including mAP and IoU. Specifically, the model excels in small object detection, which is critical in complex wildfire scenarios, and maintains higher accuracy at stringent IoU thresholds, reflecting its robustness in challenging environments. The practical implications of this work are noteworthy. The lightweight architecture of the proposed model ensures efficient resource utilization, allowing UAVs to operate for extended durations without compromising on real-time detection capabilities. Its ability to identify fires and smoke even under adverse conditions, such as occlusions and variable lighting, highlights its utility in real-world firefighting applications. Integrating drones equipped with this model into disaster management systems can significantly enhance response times, reduce damage, and improve resource allocation. Moreover, this model's capability to adapt to various scenarios makes it an indispensable tool for both wildfire prevention and emergency response. While the model showcases substantial strengths, it faces challenges when deployed under suboptimal conditions, such as low visibility caused by dense smoke or extreme lighting variations. Incorporating additional data, particularly from multi-spectral imaging technologies like thermal cameras, could mitigate these challenges by enabling the detection of heat signatures and subtle fire indicators. Such enhancements would allow the model to perform effectively even in the early stages of fire development or in obscured environments. The broader implications of this research extend beyond wildfire detection. The methodology demonstrates scalability to other critical areas, including urban fire monitoring, industrial hazard detection, and environmental pollution assessment. By adapting the model to these contexts, it can contribute significantly to various domains of disaster management and environmental monitoring. Future work should focus on addressing the current limitations to enhance the model's robustness and efficiency further. Expanding the training datasets to include diverse environmental conditions and fire scenarios will improve generalization. Moreover, optimizing computational efficiency through advanced neural network compression techniques can enhance the UAVs operational range and energy efficiency. Integrating edge computing for onboard data processing would reduce reliance on external systems, enabling faster decision-making and greater UAVs autonomy. The findings presented in this study contribute significantly to the field of UAV-assisted remote sensing, advancing the frontier of real-time disaster management technologies. By improving both detection accuracy and computational efficiency, this work lays the foundation for responsive and reliable UAVs systems capable of mitigating the devastating effects of wildfires and other natural disasters. These advancements have the potential to revolutionize emergency response strategies, ensuring better preparedness and reduced impact from environmental and disaster-related challenges.

6. Limitations and Future Directions

This study advances UAVs wildfire detection but acknowledges limitations in real-world applicability. The modified Miti-DETR model, while effective under controlled conditions, may face challenges such as variable lighting and smoke density, which can compromise detection reliability. Additionally, the model's dependence on high-quality, labeled datasets restricts its generalization across varied fire scenarios. Computational optimizations also limit UAVs operational duration and range due to significant energy demands necessary for data processing and transmission, impacting prolonged monitoring capabilities in extensive wildfire scenarios.

Future research should prioritize enhancing the model's robustness against diverse environmental conditions by incorporating a broader spectrum of training datasets, including those with varying fire behaviors. Developing unsupervised or semi-supervised learning algorithms could lessen the reliance on heavily labeled datasets. Optimizing computational efficiency through advanced compression techniques and more efficient neural architectures will extend UAVs operational capabilities. Integrating edge computing would allow more processing onboard the drones, conserving energy by reducing data transmission needs. Furthermore, including multi-spectral imaging technologies could detect heat signatures and gaseous emissions early, even before they become visible. Expanding the model to monitor other environmental and disaster scenarios, such as flooding or pollution, could also increase its utility.

By addressing these limitations and expanding research directions, we can significantly enhance the scalability and effectiveness of UAV-based systems, pushing the boundaries of what UAVs can accomplish in environmental and disaster management. This could lead to substantial societal benefits, not only in wildfire detection but across a range of applications.

7. Conclusions

In this study, we have demonstrated the effectiveness of a modified Miti-DETR model for UAV-based wildfire detection, significantly enhancing computational efficiency and detection accuracy through a redesigned AlexNet backbone and novel residual depthwise separable convolution blocks. Our introduction of a residual self-attention mechanism has successfully addressed convergence challenges in deep transformer networks, ensuring robust feature representation across complex aerial images. This design significantly reduces computational load, making it suitable for the hardware constraints of UAVs while maintaining high-quality feature extraction capabilities. Our findings show that the adapted model surpasses traditional fire detection systems in accuracy and processing speed, effectively handling the variability of real-world wildfire scenarios. The use of the FLAME dataset for validation underscores our model's versatility across diverse environmental conditions. The practical implications of our research are significant, enabling UAVs equipped with our modified Miti-DETR model to perform early fire detection and monitoring more effectively, potentially mitigating the impact of wildfires through faster response times. The proposed model demonstrated superior performance over traditional detection and SOTA models. The integration of a novel residual self-attention mechanism effectively addressed convergence issues in transformer networks, resulting in faster training and improved detection accuracy, particularly for small objects in complex aerial imagery. Additionally, the advancements we have made contribute broadly to the fields of machine learning and remote sensing, with potential applications extending to environmental monitoring, disaster management, and security surveillance. The model demonstrates superior performance compared to existing systems, particularly in small object detection and complex scenarios, as validated by higher mAP and IoU metrics.

By overcoming hardware limitations and ensuring robust detection under variable environmental conditions, the model is highly applicable to real-world disaster management scenarios. Its scalability extends beyond wildfire detection to other environmental monitoring tasks, such as flood and pollution detection, highlighting its potential for broader impact. Future research will focus on enhancing the model's robustness by integrating multi-spectral imaging, optimizing energy efficiency, and exploring semi-supervised learning methods. These advancements will ensure the model's adaptability to diverse conditions, making it an invaluable tool for disaster management and environmental monitoring.

Author Contributions: Methodology, S.M. (Shakhnoza Muksimova), S.U., S.M. (Sevara Mardieva), M.A. and Y.I.C.; Software, S.M. (Shakhnoza Muksimova), S.U. and S.M. (Sevara Mardieva); Validation, M.A. and Y.I.C.; Formal analysis, S.M. (Shakhnoza Muksimova), S.U., S.M. (Sevara Mardieva), M.A. and Y.I.C.; Resources M.A. and Y.I.C.; Data curation, M.A. and Y.I.C.; Writing—original draft, S.M. (Shakhnoza Muksimova), S.U., S.M. (Sevara Mardieva), M.A. and Y.I.C.; Writing—review and editing, S.M. (Shakhnoza Muksimova), S.U., S.M. (Sevara Mardieva), M.A. and Y.I.C.; Supervision, Y.I.C.; Project administration, M.A. and Y.I.C. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is supported by the Korean Agency for Technology and Standard under the Ministry of Trade, Industry and Energy in 2023, Project numbers are 1415181629 (Development of International Standard Technologies based on AI Model Lightweighting Technologies), 1415180835 (Development of International Standard Technologies based on AI Learning and Inference Technologies) and 1415181638 (Establishment of standardization basis for BCI and AI Interoperability).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All used datasets are available online with open access.

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

References

1. El Debeiki, M.; Al-Rubaye, S.; Perrusquía, A.; Conrad, C.; Flores-Campos, J.A. An Advanced Path Planning and UAV Relay System: Enhancing Connectivity in Rural Environments. *Future Internet* **2024**, *16*, 89. [CrossRef]
2. So, W. Forest fire cases South Korea 2011–2023. Statistics Korea. *Statista*. 2024. South Korea. Available online: www.statista.com/statistics/1296399/south-korea-forest-fire-outbreaks/ (accessed on 2 September 2024).
3. Gruffeille, C.; Perrusquía, A.; Tsourdos, A.; Guo, W. Disaster Area Coverage Optimisation Using Reinforcement Learning. In Proceedings of the 2024 International Conference on Unmanned Aircraft Systems, ICUAS, Chania, Greece, 4–7 June 2024.
4. Lakhan, A.; Elhoseny, M.; Mohammed, M.A.; Jaber, M.M. SFDWA: Secure and Fault-Tolerant Aware Delay Optimal Workload Assignment Schemes in Edge Computing for Internet of Drone Things Applications. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 1–11. [CrossRef]
5. Shakhnoza, M.; Sabina, U.; Sevara, M.; Cho, Y.I. Novel Video Surveillance-Based Fire and Smoke Classification Using Attentional Feature Map in Capsule Networks. *Sensors* **2022**, *22*, 98. [CrossRef]
6. 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]
7. Zhang, D.; Yuan, J.; Meng, H.; Wang, W.; He, R.; Li, S. Extrinsic calibration method for integrating infrared thermal imaging camera and 3D LiDAR. *Sens. Rev.* **2024**, *44*, 490–504. [CrossRef]
8. Shi, P.; Lu, J.; Wang, Q.; Zhang, Y.; Kuang, L.; Kan, X. An Efficient Forest Fire Detection Algorithm Using Improved YOLOv5. *Forests* **2023**, *14*, 2440. [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. Perrusquía, A.; Guo, W.; Fraser, B.; Wei, Z. Uncovering Drone Intentions Using Control Physics Informed Machine Learning. *Commun. Eng.* **2024**, *3*, 36. [CrossRef]
11. Yandouzi, M.; Grari, M.; Idrissi, I.; Moussaoui, O.; Azizi, M.; Ghoumid, K.; Elmiad, A.K. Review on Forest Fires Detection and Prediction Using Deep Learning and Drones. *J. Theor. Appl. Inf. Technol.* **2022**, *100*, 4565–4576.
12. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In *European Conference on Computer Vision (ECCV)*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 213–229.
13. Yuan, C.; Zhang, Y.; Liu, Z. A Survey on Technologies for Automatic Forest Fire Monitoring, Detection, and Fighting Using Unmanned Aerial Vehicles and Remote Sensing Techniques. *Can. J. For. Res.* **2015**, *45*, 783–792. [CrossRef]
14. Zhang, X.; Li, L.; Di, D.; Wang, J.; Chen, G.; Jing, W.; Emam, M. SERNet: Squeeze and Excitation Residual Network for Semantic Segmentation of High-Resolution Remote Sensing Images. *Remote Sens.* **2022**, *14*, 4770. [CrossRef]
15. Xu, Y.; Luo, W.; Hu, A.; Xie, Z.; Xie, X.; Tao, L. TE-SAGAN: An Improved Generative Adversarial Network for Remote Sensing Super-Resolution Images. *Remote Sens.* **2022**, *14*, 2425. [CrossRef]
16. Diwan, T.; Anirudh, G.; Tembhurne, J.V. Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications. *Multimed. Tools Appl.* **2023**, *82*, 9243–9275. [CrossRef] [PubMed]
17. Safarov, F.; Muksimova, S.; Kamoliddin, M.; Cho, Y.I. Fire and Smoke Detection in Complex Environments. *Fire* **2024**, *7*, 389. [CrossRef]
18. Talaat, F.M.; ZainEldin, H. An Improved Fire Detection Approach Based on YOLO-v8 for Smart Cities. *Neural Comput. Appl.* **2023**, *35*, 20939–20954. [CrossRef]
19. Abdusalomov, A.B.; Islam, B.M.S.; Nasimov, R.; Mukhiddinov, M.; Whangbo, T.K. An Improved Forest Fire Detection Method Based on the Detectron2 Model and a Deep Learning Approach. *Sensors* **2023**, *23*, 1512. [CrossRef]
20. Li, J.; Xu, R.; Liu, Y. An Improved Forest Fire and Smoke Detection Model Based on YOLOv5. *Forests* **2023**, *14*, 833. [CrossRef]
21. Lu, K.; Xu, R.; Li, J.; Lv, Y.; Lin, H.; Liu, Y. A Vision-Based Detection and Spatial Localization Scheme for Forest Fire Inspection from UAV. *Forests* **2022**, *13*, 383. [CrossRef]
22. Sudhakar, S.; Vijayakumar, V.; Kumar, C.S.; Priya, V.; Ravi, L.; Subramaniyaswamy, V. Unmanned Aerial Vehicle (UAV) Based Forest Fire Detection and Monitoring for Reducing False Alarms in Forest-Fires. *Comput. Commun.* **2020**, *149*, 1–16. [CrossRef]
23. Wahyono; Harjoko, A.; Dharmawan, A.; Adhinata, F.D.; Kosala, G.; Jo, K.-H. Real-time Forest Fire Detection Framework Based on Artificial Intelligence Using Color Probability Model and Motion Feature Analysis. *Fire* **2022**, *5*, 23. [CrossRef]

24. Khan, A.; Hassan, B.; Khan, S.; Ahmed, R.; Abuassba, A. DeepFire: A Novel Dataset and Deep Transfer Learning Benchmark for Forest Fire Detection. *Mob. Inf. Syst.* **2022**, *2022*, 5358359. [CrossRef]
25. Chopde, A.; Magon, A.; Bhatkar, S. Forest Fire Detection and Prediction from Image Processing Using RCNN. In Proceedings of the 7th World Congress on Civil, Structural, and Environmental Engineering, Virtual, 10–12 April 2022.
26. Twidwell, D.; Allen, C.R.; Detweiler, C.; Higgins, J.; Laney, C.; Elbaum, S. Smokey Comes of Age: Unmanned Aerial Systems for Fire Management. *Front. Ecol. Environ.* **2016**, *14*, 333–339. [CrossRef]
27. Skeele, R.C.; Hollinger, G.A. Aerial Vehicle Path Planning for Monitoring Wildfire Frontiers. In *Field and Service Robotics*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 455–467.
28. Wesly, U.J.; Venkata Sai Chaitanya, R.; Lakshman Kumar, P.; Siva Kumar, N.; Devi, K. A Detailed Investigation on Forest Monitoring System for Wildfire Using IoT. In Proceedings of the First International Conference on Recent Developments in Electronics and Communication Systems (RDECS-2022), Surampalem, India, 22–23 July 2022; Volume 32, p. 315.
29. Beachly, E.; Detweiler, C.; Elbaum, S.; Duncan, B.; Hildebrandt, C.; Twidwell, D.; Allen, C. Fire-aware Planning of Aerial Trajectories and Ignitions. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 685–692.
30. Kasyap, V.L.; Sumathi, D.; Alluri, K.; Ch, P.R.; Thilakarathne, N.; Shafi, R.M. Early Detection of Forest Fire Using Mixed Learning Techniques and UAV. *Comput. Intell. Neurosci.* **2022**, *2022*, 3170244. [CrossRef] [PubMed]
31. Blanco-Rodríguez, M.Á.; Ameztegui, A.; Gelabert, P.; Rodrigues, M.; Coll, L. Short-term recovery of post-fire vegetation is primarily limited by drought in Mediterranean forest ecosystems. *Fire Ecol.* **2023**, *19*, 68. [CrossRef]
32. Lucas-Borja, M.E.; Plaza-Álvarez, P.A.; Yáñez, M.D.C.; Miralles, I.; Ortega, R.; Soria, R.; Candel-Pérez, D.; Zema, D.A. Long-term evaluation of soil functionality in Mediterranean forests after a wildfire and post-fire hillslope stabilisation. *For. Ecol. Manag.* **2024**, *555*, 121715. [CrossRef]
33. Zhou, C. A2D2: AI-Driven Autonomous Drone-Based Detection Network for Wildfires. In Proceedings of the 2024 International Conference on Advanced Robotics and Intelligent Systems (ARIS), Taipei, Taiwan, 22–24 August 2024; pp. 1–6. [CrossRef]
34. Ma, W.; Zhang, T.; Wang, G. Miti-DETR: Object Detection Based on Transformers with Mitigatory Self-Attention Convergence. *arXiv* **2021**, arXiv:2112.13310.
35. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
36. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on Machine Learning (ICML), Haifa, Israel, 21–24 June 2010; pp. 807–814.
37. Shamsoshoara, A.; Afghah, F.; Razi, A.; Zheng, L.; Fulé, P.Z.; Blasch, E. Aerial Imagery Pile Burn Detection Using Deep Learning: The FLAME Dataset. *Remote Sens.* **2021**, *13*, 3671. [CrossRef]
38. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. *arXiv* **2015**, arXiv:1412.6980.