Deep Learning-based Wildfire Smoke Detection using Uncrewed Aircraft System Imagery

Khan Raqib Mahmud[1], Lingxiao Wang[2], Xiyuan Liu[3], Jiahao Li[1], and Sunzid Hassan[1]

***Abstract*— Recent years have seen notable advancements in wildfire smoke detection, particularly in Uncrewed Aircraft Systems (UAS)-based detection employing diverse deep learning (DL) approaches. Despite the promise exhibited by these approaches, the task of detecting smoke from UAS imagery remains challenging due to difficulties in differentiating smoke from similar phenomena such as clouds and water. This work introduces a novel DL-based method for smoke detection from UAS visual observations. The core idea involves segregating forest areas from non-forest regions, such as sky and lake, and exclusively applying smoke detection to forested areas, thus eliminating the chance of misidentifying clouds and water as smoke. Specifically, we utilized a Mask Region-Based Convolutional Neural Network (Mask R-CNN) for semantic segmentation to remove non-forest regions (e.g., sky and lake): Subsequently, a customized You Only Look Once-version 7 (YOLOv7) model was trained to detect smoke within the forest areas. The proposed method was validated on an image dataset collected from our previous prescribed burn experiment, where we extracted 246 images to train both MASK R-CNN and YOLOv7 models. Additionally, we extract another 128 images to validate and confirm the efficacy of our enhanced wildfire smoke detection approach. The test results demonstrate that our proposed approach, employing MASK R-CNN and YOLOv7 models, outperforms the YOLOv7-only model by 25.3% in precision, 18.7% in recall, and 45% in mean Average Precision (mAP).**

I. INTRODUCTION

Wildfires pose significant threats to ecosystems, human life, and economies, intensified by climate change, emphasizing the critical need for timely detection and control methods. Traditional wildfire detection methods, relying on remote sensors like gas, smoke, temperature, and flame detectors, have limitations in delayed response and limited coverage [1]. Thanks to recent advancement of computer vision techniques, such as object detection [2] and image segmentation [3], the detection of wildfires based on visual features becomes an effective and efficient option. Moreover, with the recent advancement in robotics, employing uncrewed aircraft systems (UAS) in wildfire detection becomes a feasible and cost-effective alternative to traditional manned aircraft surveys. Computer vision techniques, particularly smoke detection using UASs, emerge as crucial tools in addressing the challenges posed by wildfires, mitigating their environmental and societal impacts.

[1]Khan Raqib Mahmud, Jiahao Li, and Sunzid Hassan are with Computer Science Department, Louisiana Tech University, Ruston, LA 71272, USA krm070@latech.edu, jli018@latech.edu, sha040@latech.edu

[2]Lingxiao Wang is with Electrical Engineering Department, Louisiana Tech University, Ruston, LA 71272, USA lwang@latech.edu

[3]Xiyuan Liu is with Mathematics and Statistics Department, Louisiana Tech University, Ruston, LA 71272, USA liuxyuan@latech.edu

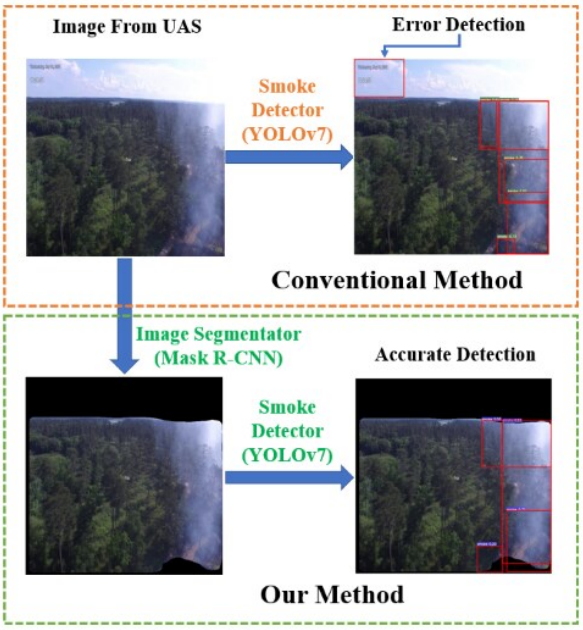


Fig. 1. An overview of smoke detection using conventional method and our proposed method. Conventional Method detects smoke using smoke detector (e.g., YOLOv7), but this method produce high error detection rate as we can see it detects sky as smoke. In our method, we first separate the forest area from the non-forest areas (e.g., sky and lake) using the Image Segmentator (e.g., Mask R-CNN) and then apply the Smoke Detector.

A common setting in computer vision-based wildfire detection is designing a deep learning (DL) model to detect smokes in images captured from the UAS's onboard cameras [3], [4]. A primary challenge is the inherent similarities between smoke and various background elements such as clouds, sky, lake and sunlight [5]. To address this research gap, we propose a two-stage DL-based approach for UAS-based smoke detection. This approach effectively identifies non-forest regions, such as sky and lake areas, and eliminates them from the images. The primary objective is to minimize the detection error rates and mitigate false alarms associated with mis-detection in the presence of complex environmental backgrounds.

The proposed methods involves two steps: semantic segmentation and smoke detection. First, we developed a DL model to perform semantic segmentation on images observed from the UAS's onboard camera to separate forest areas. We focus on the forest area since wildfire smoke comes from the ground (forest), thus, sky and lake will be considered as noisy background. Then, we designed a smoke detector to automatically sense smokes in forest areas,

eliminating the possibility of mismatching clouds as smokes. By doing so, we intend to enhance the overall accuracy and reliability of UAS-based smoke detection, contributing to more effective wildfire management strategies. Figure 1 provides an overview comparing wildfire smoke detection using the You Only Look Once-version 7 (YOLOv7) only method (conventional method) and our proposed method, which combines Mask Region-Based Convolutional Neural Network (Mask R-CNN) and YOLOv7 models.

In summary, the primary contributions of this paper are outlined as follows:

1) We propose an improved approach for detecting wildfire smoke, utilizing both the Mask R-CNN and YOLOv7 models. In this approach, the mask R-CNN effectively eliminates the non-forest regions from the images by segmenting them.
2) We collected an image dataset containing UAS wildfire images for DL training and applied various data augmentation technqiues to facilitate the trianing process.
3) We compared the proposed method with conventional smoke detection method to validate the effectiveness of our proposed method. For this purpose, we curated a new test dataset comprised of wildfire smoke images sourced from UAS imagery. This dataset predominantly includes images showcasing non-forest regions, specifically the sky and lake areas.

The remaining of the paper is organized as follows: Section II provides the current research on deep learning techniques for smoke detection and segmentation. In section III, we discuss the model preliminaries and the architectures of Mask R-CNN and YOLOv7 models. In section IV, we present an overall methodology which is capable of detecting forest fire smoke from UAS imagery and the proposed method that combines the Mask R-CNN [6] and YOLOv7 [7] for detecting the smoke from UAS images with more accuracy. Finally, in section V, we examine the experimental results and compare the accuracy of of our proposed approach with the YOLOv7 model for forest fire detection.

II. RELATED WORKS

A. Smoke Detection Techniques

Smoke detection in forest fires presents a significant challenge due to the intricate and dynamic nature of the background, compounded by environmental factors like fog, rain, and varying lighting conditions. This complexity makes deep learning (DL) techniques for wildfire smoke detection demanding. Recent research has focused on innovative approaches. Zhang et al. [8] introduced Faster R-CNN for smoke detection, eliminating manual feature extraction. Wu et al. [9] compared object detection models, highlighting SSD for real-time and accurate early fire detection. Saponara et al. [10] demonstrated improved smoke detection using YOLOv2.

In the context of Uncrewed Aircraft Systems (UASs), Jiao et al. [11] employed YOLOv3 for forest fire detection, while Jeong et al. [12] proposed a YOLOv3-LSTM hybrid. Peng et al. [13] enhanced smoke detection using SqueezeNet, showing superior speed and accuracy. Models with pruning, reconstruction, clustering [14], generalization [15], and color-motion features [16] further improve efficiency.

To address overlapping spectral signatures, Mukhiddinov et al. [5] modified YOLOv5, excelling in small smoke region detection. Kim et al. [17] introduced YOLOv7, effective but challenged by complex backgrounds. Xu et al. [18] used deep saliency networks, while Wang et al. [19] combined SSD with ViBe for video smoke detection. Jia et al. [20] and Choi et al. [21] integrated domain knowledge and SlowFast models, respectively, for enhanced detection.

These advancements highlight the diverse strategies employed to address the challenges in wildfire smoke detection using DL methods.

B. Smoke Segmentation Technques

Smoke, with its translucent and irregular characteristics, poses a complex challenge in image segmentation due to its intricate blending with the background. Small or thin smoke adds to the difficulty of accurately segmenting smoke from images. Existing segmentation algorithms struggle with capturing the variable size and translucent nature of smoke, resulting in challenges like missed or incorrect segmentation at the edges.

To address these challenges, Yuan et al. [22] proposed a deep smoke segmentation network functioning as an encoder-decoder Fully Convolutional Network (FCN) with skip structures. This network demonstrates superior differentiation between fog, clouds, and smoke, minimizing misclassifications of fog and cloud pixels as smoke. Another study introduced the Classification-assisted Gated Recurrent Network (CGRNet) [23], incorporating an Attention Convolutional GRU module (Att-ConvGRU) to extract distinguishable features, especially in scenarios with small, semi-transparent, or inconspicuous smoke.

Sun et al. [24] proposed a semi-supervised learning-based fire instance segmentation method addressing challenges related to limited labeled datasets. Smoke–U-Net [25], a multi-scale semantic segmentation approach, utilizes Multi-Scale Residual Group Attention (MRGA) in conjunction with U-Net and an encoder Transformer to extract multi-scale smoke features, notably improving the perception of small-scale smoke.

Jia et al. [26] introduced a cGAN-based model for automatic smoke region segmentation in successive video frames, demonstrating superior speed compared to saliency-based methods. However, segmentation success is highly dependent on the dataset. Convolutional Neural Network-based architectures [27] have shown efficiency in real-time segmentation for smoke and fire detection, surpassing other architectures. VSSNet [28], a 3D convolutional neural network, enhances segmentation accuracy and reduces false positives in complex natural scenes.

C. Smoke Detection and Segmentation Techniques

Accurately isolating smoke in single images is a challenging task, given the inconspicuous nature of small smoke, complex textures resulting from blending with diverse backgrounds, the multi-scale nature of evolving smoke, and interference from smoke-like objects such as haze and clouds.

Khan et al. [29] proposed an efficient smoke detection and semantic segmentation framework for clear and hazy outdoor environments. The method combines a pretrained EfficientNet for smoke detection and DeepLabv3 CNN for semantic segmentation. While showing satisfactory real imagery results, the lack of quantitative assessment raises concerns about generalizability to diverse wildfire scenarios. Ghali et al. [30] introduced a deep ensemble learning method, combining various deep convolutional models and vision transformers to address background complexity and small wildfire areas.

Xiong et al. [31] presented an SVM-based semantic segmentation using a superpixel merging algorithm to efficiently distinguish smoke from other elements, especially clouds. Wu et al. [32] proposed a sparse representation-based method for video-based smoke classification and detection. Cao et al. [33] introduced EFFNet, an enhanced feature foreground network for smoke analysis in videos, emphasizing not only smoke detection but also locating the source of smoke through semantic segmentation.

Compared to exisitng smoke detection and segmentation methods, our proposed method is novel since we combine the benefits from smoke detection and segmentation. Unlike traditional methods, we adopt a two-step process. Firstly, we employ an image segmentation model to distinguish forest areas from non-forest regions, such as the sky and lake, effectively eliminating objects with similarities to smoke. Subsequently, the smoke detection model is applied exclusively to images containing forested areas, minimizing the chances of error detection in non-forest regions. This novel strategy enhances the accuracy and reliability of our smoke detection methodology.

III. Preliminaries of Mask R-CNN and YOLOv7

Since this work is built upon two computer vision models, i.e., Mask R-CNN and YOLOv7. It is necessary to introduce these two models before introducing our proposed method.

A. Mask R-CNN Architecture

Mask R-CNN [6] is a deep learning (DL) model designed for image segmentation task. The architecture of the Mask R-CNN model, can be divided into two primary components: 1) The convolutional backbone, which conducts feature extraction across the entire image; 2) The network head, which handles individual aspects of bounding-box recognition (i.e., classification and regression) and mask prediction for each Region of Interest (RoI).

Fig. 2 presents the architecture of Mask R-CNN [6]. Initially, a pre-trained CNN like ResNet [35] serves as the backbone to process the input image and capture essential features, generating Feature Maps. Following this, the Region Proposal Network (RPN) operates on the backbone's feature map, suggesting potential regions or bounding boxes that may contain objects in the image. After RPN generates these proposals, the ROIAlign (Region of Interest Align) layer is introduced to address alignment issues and precisely extract features for accurate pixel-wise segmentation.

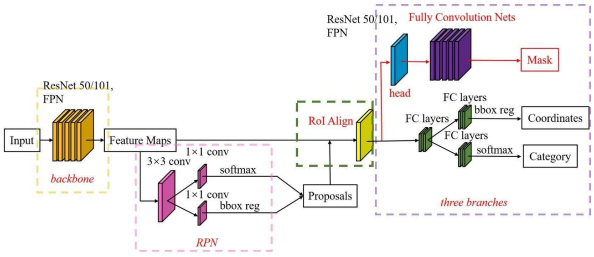


Fig. 2. Network Architecture of Mask R-CNN. The image is retrieved from [34].

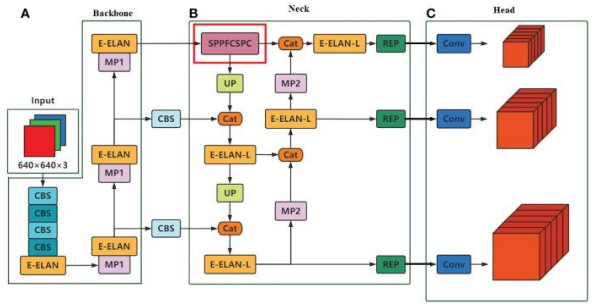


Fig. 3. Network Architecture of YOLOv7. The image is retrieved from [37].

The outputs of Mask R-CNN include three components, a binary mask for object segmentation, bounding boxes of detected objects in the image, and the categories of the objects. Specifically, a mask head is responsible for creating segmentation masks for each proposed region. By utilizing the features aligned through ROIAlign, the mask head predicts binary masks, outlining pixel boundaries for each object. Other two outputs, i.e., bounding box locations and object categories, are computed via two branches with fully-connected (FC) layers.

B. YOLOv7 Architecture

YOLOv7 [7], the latest variant in the YOLO model family [36], introduces key advancements. It incorporates the Extended Efficient Layer Aggregation Network (E-ELAN) to efficiently manage short and long gradient paths, enhancing learning capabilities. Additionally, YOLOv7 addresses model scaling for concatenation-based models, tailoring characteristics for diverse applications.

The YOLOv7 architecture, illustrated in Fig. 3, comprises backbone, neck, and head components. The backbone generates diverse feature maps, preserving multi-scale information. The neck merges these maps, incorporating fine-grained details and deep semantic information. The head transforms integrated features into detection predictions.

The backbone extracts features using four CBS modules, including Convolution, Batch normalization, and SiLU activation function. The E-ELAN and MP modules sequentially extract features, resulting in three E-ELAN modules that feed into the neck. The MP module combines MaxPool and CBS, while the E-ELAN module consists of multiple convolutional layers. Outputs from E-ELAN modules are input to the neck network.

The neck, adopting a Path Aggregation Feature Pyramid Network (PAFPN) structure, combines elements from FPN [38] and PANet [39]. The SPP structure extracts attributes at multiple scales, and the CSP design results in the SPPFCSPC structure, expanding the network's perception. Rep modules, adjusting channel numbers, follow PAFPN.

Sequential feature extraction concludes with 1×1 convolutions, integrating features for a seamless transition to actionable predictions within the YOLOv7 framework.

IV. IMPROVED WILDFIRE SMOKE DETECTION USING MASK R-CNN AND YOLOV7

This study aims to enhance the precision of wildfire smoke prediction through the integration of the YOLOv7 model using UAS imagery. As detailed in related studies (Section II), detecting wildfire smoke remains challenging due to its similarity to various background elements, such as sky and lake regions, along with varying lighting conditions. Moreover, accurate smoke predictions using DL methods are hindered by the limited availability of labeled image dataset for wildfire smokes. To address these challenges, we propose an improved wildfire smoke detection approach that leverages both mask R-CNN and YOLOv7.

A. An Overview of the Proposed Methods

The proposed methodology, illustrated in Fig. 4, follows a systematic approach to enhance forest fire smoke detection. Initially, a Forest Segmentator, represented by a Mask R-CNN model, is employed to conduct image segmentation, effectively extracting non-forest regions, such as sky and lake areas. Subsequently, a binary mask is generated based on the segmented image, excluding the identified Non-Forest regions. Finally, a Forest Smoke Detector, utilizing the YOLOv7 model, is applied to detect smoke specifically within the masked image, which now contains only forest regions. This sequential process ensures a focused and accurate detection of smoke specifically within the forested areas of the images.

B. Smoke Detection using YOLOv7 Model

In our methodology, we fine-tuned the YOLOv7 model using a dataset specifically collected from UAS imagery, focusing on forest fire smoke images for smoke detection. The initial model was pre-trained on the MSCOCO dataset [7]. Subsequently, we trained our model using the collected dataset. To evaluate the model's performance, we utilized an unseen test dataset. During this assessment, we observed

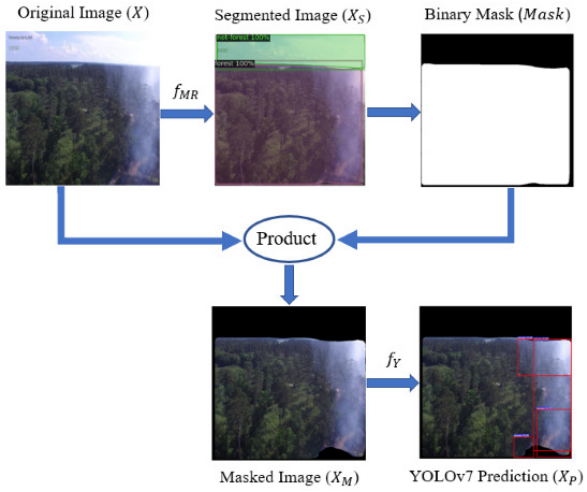


Fig. 4. Overall framework of our proposed method. In our approach, the original image denoted as X serves as the input to the Image Segmentator function, represented by f_{MR}. The resulting output, denoted as X_S, represents the Segmented Image. A Binary Mask, designated as $Mask$, is generated based on this Segmented Image (X_S). The Masked Image, denoted as X_M, is then obtained through element-wise multiplication between the Original Image (X) and the Binary Mask ($Mask$). Finally, the Masked Image (X_M) is fed into the Smoke Detector function, denoted as f_Y, providing the ultimate prediction for smoke detection.

instances of misidentification, notably in images with backgrounds such as sky and lake regions. These areas, predominantly categorized as non-forest regions within the wildfire smoke dataset, posed challenges for accurate identification.

C. Mask Generation and Smoke Detection

To address misdetection in non-forest regions, our strategy involves creating a binary mask for exclusion. Using a Forest Segmentator, implemented by a Mask R-CNN model trained on the same dataset as the YOLOv7 model, we segmented the input images into forest and non-forest regions, including the sky and lake areas. The segmentation process is represented as follows:

$$X_S = f_{MR}(X) \tag{1}$$

where f_{MR} is the Image Segmentator, X is the imput image and X_S is the segmented image, as Fig. 4 illustrated.

Subsequently, a binary mask corresponding to these forest and non-forest regions was generated from the segmented image, X_S to create the masked image:

$$Mask(i,j) = \begin{cases} 1 & \text{if } (i,j) \in Forest \\ 0 & \text{if } (i,j) \in Non-Forest \end{cases} \tag{2}$$

where $Mask$ is the binary mask, and (i,j) represents the position of pixel values in the segmented image, X_S, as depicted in Fig. 4.

The masked image of the original image was generated by applying the binary mask to the image, effectively eliminating the non-forest areas and minimizing the chances of smoke misdetection in those regions:

$$X_M = X * Mask \tag{3}$$

where X_M is the masked image of the original image and $(*)$ represents element-wise multiplication.

The resulting masked image, containing only forest regions, was then fed into the Forest Smoke Detector, represented by the YOLOv7 model, for smoke detection:

$$X_P = f_Y(X_M) \tag{4}$$

where f_Y is the Forest Smoke Detector and X_P is the predicted image for smoke detection, as illustrated in Fig. 4.

Notice that we can express the expected value of the smoke detector (e.g., YOLOv7) for the original image, X, by the total sum of the probabilities. That is,

$$\begin{aligned} E(f_Y(X)) &= P(smoke) \\ &= P(smoke \cap forest) + \\ &\quad P(smoke \cap non-forest) \end{aligned} \tag{5}$$

Furthermore, given the masked image, X_M, we have,

$$\begin{aligned} E(f_Y(X_M)) &= P(smoke \cap forest) \\ &= P(smoke|forest) \cdot P(forest) \\ &\leq E(f_Y(X)) \end{aligned} \tag{6}$$

where $E(f_Y(X))$ and $E(f_Y(X_M))$ are the expected values of smoke detector for X and X_M respectively. $P(smoke)$ represents the probability that the smoke detector detects the smoke, and $P(forest)$ represents the probability that the area is forest. Hence, $P(smoke|forest)$ is the probability that the smoke detector detects smoke, given that this is a forest area.

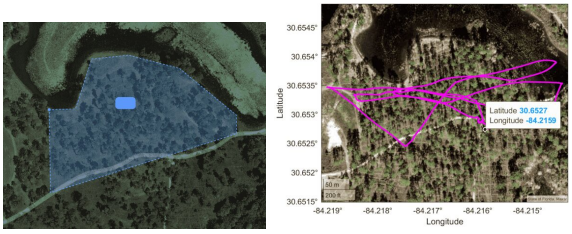
V. EXPERIMENTS AND RESULTS

This section outlines the experimental setup and the outcomes related to wildfire smoke detection, employing both the YOLOv7 model and a novel hybrid approach that integrates mask R-CNN with YOLOv7 for enhanced performance in smoke detection.

A. Video Recording from Prescribed Burn

In May 2022, we collaborated with the Tall Timber fire institution to conduct a prescribed burn at Tallahassee, Florida. Fig. 5(a) shows the satellite image of the prescribed burning area. The size of the burn area is approximately 9 acres, which is a forest land inside the Tall Timber fire institution. In the spring season, the Tall Timber fire institution conducts a prescribed burn to eliminate weeds and fertilize the land.

During the prescribed burn, we deployed a multirotor UAS and successfully collected data during the burn. In the flight, the UAS started at the downwind area of the burning region and flew upwind toward the burning region. The UAS was remotely controlled by a human operator, and the sensor data was transmitted to the ground station for live monitoring of the wildfires. Fig. 5(b) presents the trajectories of the flight, which was recorded from the onboard GPS. We can see that the UAS crossed the burning areas multiple times to collect the environmental data. The flight was conducted at the early phase of the prescribed burn, i.e., right after the ignition, and returned to the ground after around 15 minutes.



(a) (b)

Fig. 5. Deploying an UAS in a prescribed burn to collect environment data. (a) The prescribed burn area (highlighted with the blue color). (b) The flight trajectory of the UAS, where the UAS is controlled by the human operator. The labeled position in the diagram is the start position.

B. Evaluation Metrices

To evaluate the effectiveness of our models, we employ 3 key metrics, including precision, recall, and mean Average Precision (mAP). Precision (P) and recall (R) are computed using the formulas in equation 7 where True Positives (TP) denote instances where the predicted value aligns with the true value, and False Positives (FP) represent incorrect detections. Instances where the detection model fails to identify a ground truth are classified as False Negatives (FN). True Negatives (TN) occur when the detection model correctly recognizes the absence of an object in images without objects.

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}. \tag{7}$$

Analyzing inequality 6, it is evident that utilizing the mask decreases the probability of smoke detection. This is due to the decrease in the value of $TP + FP$ for $f_Y(X_M)$. This reduction leads to an enhancement in precision, as defined by equation 7. The model's average precision (AP) corresponds to the area under the curve of the Precision-Recall ($P(R)$) Curve, where higher values indicate superior classifier performance.

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

In target detection, the model typically identifies multiple classes of targets, each plotting its own PR curve and calculating an AP value. The mean Average Precision (mAP) represents the average of the APs across all classes.

$$mAP = \frac{1}{N} \sum_1^N AP,$$

where N is the number of classes in the test images.

C. Dataset Preprocessing

To generate training images, we extracted 246 frames from a UAS-recorded video during a prescribed burning experiment, sampled at one frame every 3 seconds and recorded at 30 frames per second with a resolution of 1920×1080. Designated as 'Dataset-1', this dataset was used for YOLOv7 and Mask R-CNN training for smoke detection and forest

segmentation. The images within 'Dataset-1' were resized to 640×640 pixels for consistency in the evaluation process.

Additionally, we created 'Dataset-2', comprising 128 unseen smoke images from the UAS video for testing. Emphasizing non-forest regions like sky and lake, 'Dataset-2' serves for evaluating our wildfire smoke detection approach. Roboflow [40] was utilized as the annotation tool for accurate bounding boxes and polygons delineation.

To assess YOLOv7 performance, diverse predefined augmentation techniques in Roboflow were systematically applied to 'Dataset-1'. These included rotation (-10° to +10°), shear (±15° horizontally and vertically), hue adjustment (-25° to +25°), saturation adjustment (-25% to +25%), brightness adjustment (-25% to +25%), exposure adjustment (-25% to +25%), blur (up to 2.5px), and noise (up to 1% of pixels). Post-augmentation, the resulting augmented dataset, labeled as 'Dataset-3', enriched the training set for a comprehensive evaluation of YOLOv7's robustness in detecting prescribed wildfire smoke.

D. Smoke Detection using YOLOv7

To evaluate the effectiveness of YOLOv7, we conducted experiments using various versions of pre-trained models, including standard models such as YOLOv7 and YOLOv7-W6, as well as a compound scaling model named YOLOv7-X. All these models were trained utilizing the Microsoft COCO dataset [7]. The parameter counts for these pre-trained YOLOv7 models are 36.9 million, 70.4 million, and 71.3 million for YOLOv7, YOLOv7-W6, and YOLOv7-X, respectively [7].

The training of YOLOv7 models involved consistent parameters across different pre-trained versions. We set the number of training epochs to 100, with a batch size of 16. The training utilized two distinct datasets, namely 'Dataset-1' and the augmented set 'Dataset-3', while performance evaluation was conducted on the test dataset, 'Dataset-2'. During testing, specific parameters were applied, including an IoU (Intersection over Union) threshold of 0.65 and a confidence score threshold of 0.01.

From Fig. 7 (a)-(e), we observe that by employing YOLOv7-only, instances of wildfire smoke misdetection occur and these misdetections commonly occur in images dominated by non-forest regions, particularly in sky and lake areas, as well as in lake regions illuminated by sunlight. This insight underscores the importance of robustly addressing detection challenges in diverse environmental contexts.

E. Exclusion of Non Forest Regions using Mask R-CNN

In the segmentation aspect of our experiments, we employed Detectron 2 [41] to train a pre-trained version of the Mask R-CNN model using 'Dataset-1'. This aimed to effectively segment non-forest regions, specifically sky and lake, in the images. The pre-trained Mask R-CNN model was obtained from the 'Detectron2 Model Zoo', housing official baseline models trained with Detectron2. For this purpose, we selected the COCO Instance Segmentation Baseline model with Mask R-CNN 'R50-FPN' [42], utilizing a ResNet+FPN backbone with standard convolutional and fully connected heads for mask and box prediction.

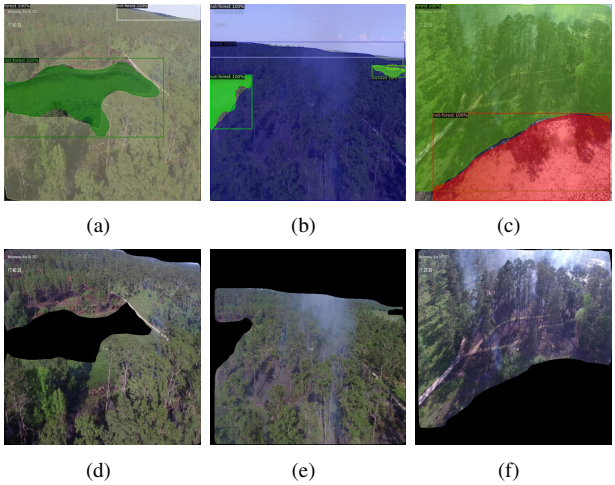


Fig. 6. Examples of semantic segmentation of 'forest' and 'non-forest' and masked images with forest regions only. (a), (b), (c) are results from semantic segmentation using Mask R-CNN. (d), (e), (f) are masked images obtained using Eqn. 3.

We employed polygons to identify the boundaries of both forest and non-forest regions in the images, labeled as 'forest' and 'non-forest', respectively. We trained the Mask R-CNN model to perform segmentation of images into forest and non-forest regions. Subsequently, the trained model was applied to the test dataset 'Dataset-2' to predict 'forest' and 'non-forest' regions in images.

The segmented images were then utilized to generate masks specifically for non-forest regions within the images. This involved setting all pixel values corresponding to non-forest regions to zero, following the process discussed in section IV-C, effectively creating a mask representing the non-forest areas within the images. This segmentation process provides a critical foundation for assessing the precision of our proposed approach in identifying and isolating forest regions affected by prescribed wildfire smoke.

In Figure 6, we illustrate instances of segmentation for non-forest regions and the subsequent creation of masks for these segmented areas. Our observations indicate a notable accuracy in effectively excluding non-forest regions from images that predominantly feature such areas, especially in the sky and lake regions, including lake areas illuminated by sunlight. This underscores the efficacy of our segmentation strategy in accurately identifying and isolating prescribed wildfire smoke-affected areas within the overall imagery.

F. Comparison of Proposed approach for smoke detection using YOLOv7

The rationale behind our approach is to eliminate non-forest regions, particularly sky and lake areas, as these regions share complex and sometimes similar characteristics with smoke. As discussed in Section V-E, we excluded the sky and lake regions from the images. Subsequently, we utilized these modified images, excluding non-forest regions,

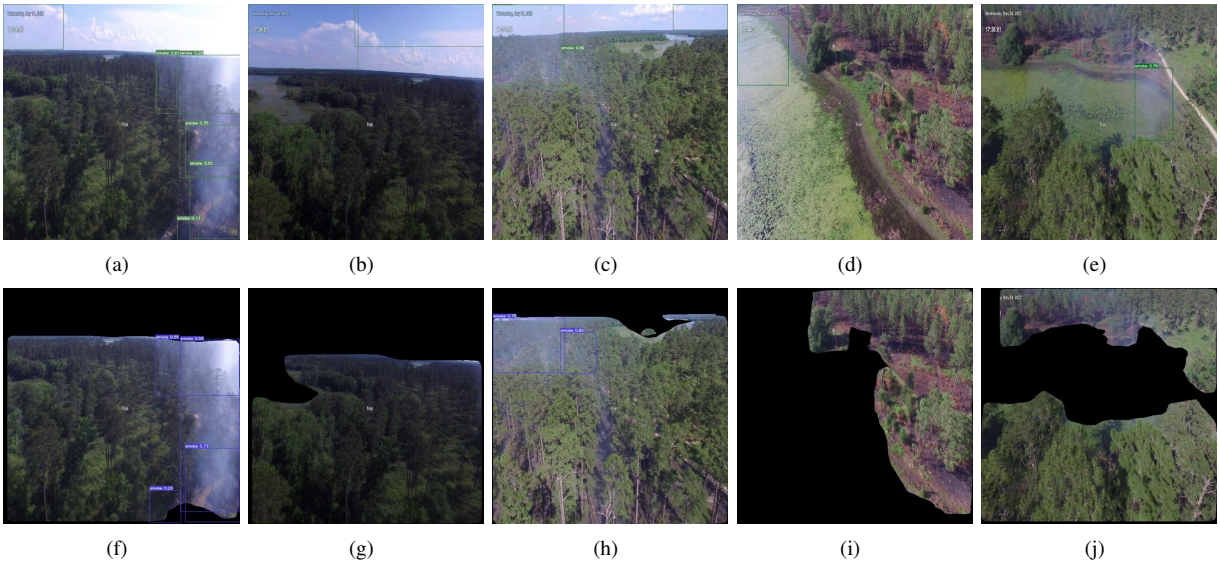


Fig. 7. Smoke Detection using conventional approach of YOLOv7-only model and our proposed approach using Mask R-CNN and YOLOv7 models. (a), (b), (c), (d), (e) are results from conventional approach. (f), (g), (h), (i), (j) are results from our approach.

TABLE I
COMPARISON OF EVALUATION METRICES BETWEEN THE YOLOV7 MODEL AND OUR PROPOSED APPROACH EMPLOYING BOTH MASK R-CNN AND YOLOV7 MODELS

Model	Precision		Recall		mAP	
	Before Mask	After Mask	Before Mask	After Mask	Before Mask	After Mask
YOLOv7	0.318	**0.398**	0.300	**0.356**	0.131	**0.190**
YOLOv7-w6	0.374	**0.407**	0.238	**0.297**	0.148	**0.158**
YOLOv7-X	0.347	**0.432**	0.305	**0.297**	0.227	**0.260**

to predict wildfire smoke using the YOLOv7 model, which constitutes our proposed methodology.

Table I presents the performance metrics of our proposed wildfire smoke detection approach, employing Mask R-CNN and YOLOv7 models, compared to the YOLOv7-only model. The evaluation metrics were computed using the test dataset, 'Dataset-2'.

Upon comparing the performance of YOLOv7 models, as outlined in Table I, we observed that our proposed method, employing Mask R-CNN and YOLOv7 models for smoke detection, achieved higher accuracy than the conventional approach using YOLOv7 only. To validate our approach, we utilized the test dataset 'Dataset-2', where a significant portion of the data was manually selected, emphasizing more examples with complex backgrounds and non-forest regions within the images.

Figure 7 provides visual examples of wildfire smoke detection using both the conventional approach of YOLOv7 model and our proposed approach employing Mask R-CNN model with YOLOv7 model. The results illustrate that excluding non-forest regions from images, especially those containing sky and lake regions (including areas illuminated by sunlight), leads to improved performance and a substantial increase in accuracy. This underscores the effectiveness of our methodology in enhancing the precision of wildfire smoke detection in challenging scenarios with diverse backgrounds.

VI. CONCLUSIONS

In our analysis, we identified challenges in YOLOv7-based forest fire smoke detection, especially in the presence of non-forest elements like sky and lake, and difficulties in capturing smoke accurately in challenging lighting conditions. To enhance precision, we propose an improved approach leveraging mask R-CNN and YOLOv7 models. Our method exhibits notable accuracy improvements, particularly in images with non-forest elements. Integrating the mask R-CNN minimizes smoke misdetection by segmenting non-forest areas, addressing challenges posed by complex backgrounds and enhancing overall precision.

To assess YOLOv7's precision, we explored diverse pre-trained versions on both original and augmented datasets, providing a comprehensive evaluation of adaptability and performance across scenarios. For validation, we curated a dataset emphasizing non-forest elements, resulting in enhanced accuracy, reinforcing our approach's effectiveness.

In addition, by analyzing equation 6, we can identify two extreme situations. If the mask doesn't cover any part of the image, the probability of forest remains at 1. Consequently, the expected value of the smoke detector, both before and after applying the mask to the image, remains unchanged, rendering the masking process ineffective. On the contrary, when the mask covers the entire image, the probability of forest becomes zero. Consequently, the expected value drops to zero, indicating the absence of smoke detection.

In conclusion, our improved methodology advances UAS-based wildfire smoke detection, addressing challenges and paving the way for more accurate and reliable strategies.

REFERENCES

[1] A. Gaur, A. Singh, A. Kumar, K. S. Kulkarni, S. Lala, K. Kapoor, V. Srivastava, A. Kumar, and S. C. Mukhopadhyay, "Fire sensing technologies: A review," *IEEE Sensors Journal*, vol. 19, no. 9, pp. 3191–3202, 2019.

[2] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digital Signal Processing*, vol. 126, p. 103514, 2022.

[3] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3523–3542, 2021.

[4] X. Chen, B. Hopkins, H. Wang, L. O'Neill, F. Afghah, A. Razi, P. Fulé, J. Coen, E. Rowell, and A. Watts, "Wildland fire detection and monitoring using a drone-collected rgb/ir image dataset," *IEEE Access*, vol. 10, pp. 121301–121317, 2022.

[5] M. Mukhiddinov, A. B. Abdusalomov, and J. Cho, "A wildfire smoke detection system using unmanned aerial vehicle images based on the optimized yolov5," *Sensors*, vol. 22, no. 23, 2022.

[6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, 2017.

[7] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7464–7475, 2023.

[8] Q. xing Zhang, G. hua Lin, Y. ming Zhang, G. Xu, and J. jun Wang, "Wildland forest fire smoke detection based on faster r-cnn using synthetic smoke images," *Procedia Engineering*, vol. 211, pp. 441–446, 2018. 2017 8th International Conference on Fire Science and Fire Protection Engineering (ICFSFPE 2017).

[9] S. Wu and L. Zhang, "Using popular object detection methods for real time forest fire detection," in *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 01, pp. 280–284, 2018.

[10] S. Saponara, A. Elhanashi, and A. Gagliardi, "Real-time video fire/smoke detection based on cnn in antifire surveillance systems," *Journal of Real-Time Image Processing*, vol. 18, no. 3, pp. 889–900, 2021.

[11] Z. Jiao, Y. Zhang, J. Xin, L. Mu, Y. Yi, H. Liu, and D. Liu, "A deep learning based forest fire detection approach using uav and yolov3," in *2019 1st International Conference on Industrial Artificial Intelligence (IAI)*, pp. 1–5, 2019.

[12] M. Jeong, M. Park, J. Nam, and B. C. Ko, "Light-weight student lstm for real-time wildfire smoke detection," *Sensors*, vol. 20, no. 19, 2020.

[13] Y. Peng and Y. Wang, "Real-time forest smoke detection using hand-designed features and deep learning," *Computers and Electronics in Agriculture*, vol. 167, p. 105029, 2019.

[14] C. Li, B. Yang, H. Ding, H. Shi, X. Jiang, and J. Sun, "Real-time video-based smoke detection with high accuracy and efficiency," *Fire Safety Journal*, vol. 117, p. 103184, 2020.

[15] H. Liu, F. Lei, C. Tong, C. Cui, and L. Wu, "Visual smoke detection based on ensemble deep cnns," *Displays*, vol. 69, p. 102020, 2021.

[16] M. Hashemzadeh and A. Zademehdi, "Fire detection for video surveillance applications using ica k-medoids-based color model and efficient spatio-temporal visual features," *Expert Systems with Applications*, vol. 130, pp. 60–78, 2019.

[17] S.-Y. Kim and A. Muminov, "Forest fire smoke detection based on deep learning approaches and unmanned aerial vehicle images," *Sensors*, vol. 23, no. 12, 2023.

[18] G. Xu, Y. Zhang, Q. Zhang, G. Lin, Z. Wang, Y. Jia, and J. Wang, "Video smoke detection based on deep saliency network," *Fire Safety Journal*, vol. 105, pp. 277–285, 2019.

[19] G. Wang, J. Li, Y. Zheng, Q. Long, and W. Gu, "Forest smoke detection based on deep learning and background modeling," in *2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, pp. 112–116, 2020.

[20] Y. Jia, W. Chen, M. Yang, L. Wang, D. Liu, and Q. Zhang, "Video smoke detection with domain knowledge and transfer learning from deep convolutional neural networks," *Optik*, vol. 240, p. 166947, 2021.

[21] M. Choi, C. Kim, and H. Oh, "A video-based SlowFastMTB model for detection of small amounts of smoke from incipient forest fires," *Journal of Computational Design and Engineering*, vol. 9, pp. 793–804, 04 2022.

[22] F. Yuan, L. Zhang, X. Xia, B. Wan, Q. Huang, and X. Li, "Deep smoke segmentation," *Neurocomputing*, vol. 357, pp. 248–260, 2019.

[23] F. Yuan, L. Zhang, X. Xia, Q. Huang, and X. Li, "A gated recurrent network with dual classification assistance for smoke semantic segmentation," *IEEE Transactions on Image Processing*, vol. 30, pp. 4409–4422, 2021.

[24] G. Sun, Y. Wen, and Y. Li, "Instance segmentation using semi-supervised learning for fire recognition," *Heliyon*, vol. 8, no. 12, p. e12375, 2022.

[25] Y. Zheng, Z. Wang, B. Xu, and Y. Niu, "Multi-scale semantic segmentation for fire smoke image based on global information and u-net," *Electronics*, vol. 11, no. 17, 2022.

[26] Y. Jia, H. Du, H. Wang, R. Yu, L. Fan, G. Xu, and Q. Zhang, "Automatic early smoke segmentation based on conditional generative adversarial networks," *Optik*, vol. 193, p. 162879, 2019.

[27] S. Frizzi, M. Bouchouicha, J.-M. Ginoux, E. Moreau, and M. Sayadi, "Convolutional neural network for smoke and fire semantic segmentation," *IET Image Processing*, vol. 15, no. 3, pp. 634–647, 2021.

[28] G. Zhu, Z. Chen, C. Liu, X. Rong, and W. He, "3d video semantic segmentation for wildfire smoke," *Machine Vision and Applications*, vol. 31, no. 50, 2020.

[29] S. Khan, K. Muhammad, T. Hussain, J. D. Ser, F. Cuzzolin, S. Bhattacharyya, Z. Akhtar, and V. H. C. de Albuquerque, "Deepsmoke: Deep learning model for smoke detection and segmentation in outdoor environments," *Expert Systems with Applications*, vol. 182, p. 115125, 2021.

[30] R. Ghali, M. A. Akhloufi, and W. S. Mseddi, "Deep learning and transformer approaches for uav-based wildfire detection and segmentation," *Sensors*, vol. 22, no. 5, 2022.

[31] D. Xiong and L. Yan, "Early smoke detection of forest fires based on svm image segmentation," *Journal of Forest Science*, vol. 65, no. 4, pp. 150–159, 2019.

[32] X. Wu, X. Lu, and H. Leung, "Video smoke separation and detection via sparse representation," *Neurocomputing*, vol. 360, pp. 61–74, 2019.

[33] Y. Cao, Q. Tang, X. Wu, and X. Lu, "Effnet: Enhanced feature foreground network for video smoke source prediction and detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 1820–1833, 2022.

[34] Jacqueline, "Mask r-cnn." `https://zhuanlan.zhihu.com/p/62492064?ref=blog.roboflow.com`, 2019.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[36] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[37] J. Yan, Z. Zhou, D. Zhou, B. Su, Z. Xuanyuan, J. Tang, Y. Lai, J. Chen, and W. Liang, "Underwater object detection algorithm based on attention mechanism and cross-stage partial fast spatial pyramidal pooling," *Front. Mar. Sci*, 2022.

[38] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[39] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8759–8768, 2018.

[40] B. Dwyer, J. Nelson, J. Solawetz, and et. al., "Roboflow (version 1.0)." `https://roboflow.com.computervision`, 2022.

[41] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2." `https://github.com/facebookresearch/detectron2`, 2019.

[42] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386–397, 2020.