Wildfire Detection Powered by Involutional Neural Network and Multitask Learning With Dark Channel Prior Technique

Md. Fahim-Ul-Islam, Noshin Tabassum, Amitabha Chakrabarty, Syed Mahfuzul Aziz, *Senior Member, IEEE*, Mahdieh Shirmohammadi, Nasim Khonsari, Hyun-Han Kwon, and Md. Jalil Piran, *Senior Member, IEEE*

***Abstract*—Wildfires are crucial to the ecosystem health, yet wildfires also pose a grave threat to human lives and the environment. Early detection remains challenging, especially in remote areas. Satellite data and autonomous aerial vehicles (AAVs) have become crucial tools for detecting and monitoring wildfires. Achieving proper identification is demanding due to the intricacy of wildfires, which are impacted by various environmental factors. Recent advances in remote sensing data analysis, especially the use of machine learning models, demonstrate captivating potential. Applying machine learning in remote sensing data analysis, this study examines state-of-the-art wildfire identification strategies based on the modified involutional neural network (Inv-Net). Inv-Net is used for wildfire segmentation and classification that enhances the U-Net architecture by incorporating the involutional neural network (INN) as its backbone. The INN layers dynamically perform kernel alterations, allowing for more efficient extraction of features and improved capture of spatial correlations. By utilizing the concept of multitask learning, Inv-Net achieves an accuracy of 98.1% in the classification task and exhibits a noteworthy Intersection over Union score of 97.6% in the segmentation task, notably when used for wildfire image analysis. To improve wildfire detection accuracy and resilience, Inv-Net uses the synergies between classification and segmentation. Furthermore, the dark channel prior (DCP) technique is used to address atmospheric disturbances such as haze and fog in static images. DCP reduces optical interference by estimating depth maps and transmission maps. The results highlight the importance of the DCP method, which achieves an excellent structural similarity index measure of 0.7831 and peak-signal-to-noise ratioof 17.4321, indicating greater image dehazing methods. In comprehensive trials and comparisons, we demonstrate the robustness of our approach in identifying and mapping wildfires in challenging conditions.**

***Index Terms*—Classification, dark channel prior, Inv-Net, involutional neural network, multitask learning, segmentation, wildfire.**

Received 29 April 2024; revised 18 July 2024; accepted 19 August 2024. Date of publication 27 August 2024; date of current version 28 October 2024. This work was supported by Cooperative Research Method and Safety Management Technology in National Disaster funded by Ministry of Interior and Safety (MOIS, Korea) under Grant 2022-MOIS63-001(RS-2022-ND641011).
(Corresponding author: Md. Jalil Piran.)

Md. Fahim-Ul-Islam, Noshin Tabassum, and Amitabha Chakrabarty are with the Department of Computer Science and Engineering, Brac University, Dhaka 1212, Bangladesh (e-mail: fahim.islam@bracu.ac.bd; ext.noshin.tabassum@bracu.ac.bd; amitabha@bracu.ac.bd).

Syed Mahfuzul Aziz is with the Brac University, Dhaka 1212, BangladeshUniSA STEM, University of South Australia, Mawson Lakes, SA 5095, Australia (e-mail: mahfuz.aziz@bracu.ac.bd).

Mahdieh Shirmohammadi is with the Institute of Artificial Intelligence, Shaoxing University, Shaoxing 312000, China (e-mail: mshirmohammadi.khu@gmail.com).

Nasim Khonsari is with the Westcliff University, College of Business, Westcliff University, Irvine, CA 92614 USA (e-mail: N.Khonsari.198@westcliff.edu).

Hyun-Han Kwon is with the Department of Civil and Environment Engineering, Sejong University, Seoul 05006, South Korea (e-mail: hkwon@sejong.ac.kr).

Md. Jalil Piran is with the Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea (e-mail: piran@sejong.ac.kr).
Digital Object Identifier 10.1109/JSTARS.2024.3450714

I. Introduction

WILDFIRES are a natural occurrence that is crucial for the Earth's ecosystem, but they also present substantial dangers to both the environment and human populations [1], [2], [3], [4]. Although wildfires promote vegetation growth and recycle nutrients in forests, they can also cause destruction. On a global scale, they contribute to global warming and pose a threat to ecosystems, infrastructure, and human life. The magnitude of the destruction is apparent in the astounding data, such as the National Interagency Fire Center's claim of more than 67 000 wildfires occurring each year in the United States alone, resulting in the consumption of extensive areas of land [5]. Between 2013 and 2022, the United States had an average of 61 410 wildfires per year, impacting around 7.2 million acres annually [5]. Since 1983, the National Interagency Fire Center has recorded an annual average of approximately 70 000 in wildfires in the United States. Over the past 20 years, California has experienced an average of 317 wildfires per year. In the last decade, there has been a global increase in the occurrence and severity of wildfires. That poses a threat to both human and animal lives, causing harm to ecosystems, and disturbing social equilibrium [6], [7]. There has been a worrisome pattern of more frequent and larger wildfires in California over the past two decades compared to the previous eight decades (1920–1999) [8]. As of June 1, 2023, there have been around 18 300 wildfires this year, which have destroyed over 511 000 acres of land [9]. These figures highlight the ongoing danger and increasing seriousness of wildfires around the globe, requiring proactive actions for prevention, mitigation, and management. As shown in Fig. 1, the number of hectares burning has increased at an alarming rate for the past 20 years [10], [11]. Besides, in Canada, wildfires are the main stand-replacing disruption regime in the boreal forests [12].

Early identification of wildfires poses various challenges, especially in far-off areas where clear vision and site monitoring

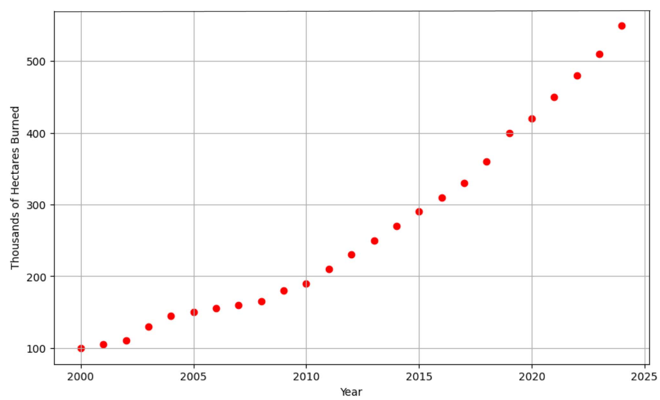

Fig. 1. Wildfire Burns Over the Years from 2000–2024 [9], [10], [11].

are challenging due to various environmental factors. To deal with this problem, remote sensing has emerged as a prominent approach for detecting wildfires, utilizing satellite data to speedily ascertain and monitor wildfires [3], [13], [14], [15]. In addition, autonomous aerial vehicles (AAVs), commonly known as drones, have become a growing significance due to their capacity to collect and transmit up-to-date information and visuals [16]. This provides a supportive resource for detecting and managing wildfires, alongside classic methodologies. These technologies are essential for improving the efficiency and effectiveness of wildfire response activities, allowing for faster intervention and mitigation methods to protect lives, property, and natural ecosystems [1], [17].

Wildfires are sophisticated natural occurrences that incorporate multiple dependent variables, such as weather conditions, land features, soil wetness, fuel composition, and the exact qualities and whereabouts of the ignition source [1], [18]. Furthermore, wildfires are distinguished by their ever-changing nature and can emerge across many spatial and temporal dimensions [19]. Therefore, the correct identification and separation of wildfires using standard remote sensing methods is frequently problematic due to the necessity of modeling intricate functional linkages among multiple data points. Cloud cover and smoke, which are produced during the burning process, also hamper the capacity to detect fires from space [20]. Accurate categorization and division of wildfires are vital for analyzing the advancement of a wildfire, ascertaining its effect on neighboring regions, and allocating resources to manage and lessen its consequences. These actions allow authorities to identify the order of importance for evacuation zones, distribute firefighting resources effectively, and examine the aftermath of flames, including soil erosion and habitat loss. Fortunately, the increased availability of remote-sensing data and the major developments in processing capacity in recent years have enhanced the possibility of applying data-centric strategies for identifying and mapping wildfires [21].

In a study by Pecha et al. [22], wildfire identification in the Alaska regions was tackled as a semantic segmentation challenge using support vector machine classifiers. The authors applied normalized transmittance spanning 152 days to represent color information, reaching high accuracy in wildfire detection. A more recent study by Akbari Asanjan et al. [23] has developed a probabilistic wildfire segmentation strategy employing a supervised deep generative model from satellite photos. This approach developed stochastic segmentation by two model streams, one learning meaningful stochastic latent distributions and the other generating high-quality segmentation masks. Besides, a study by Ghali et al. [24] has employed deep vision transformers for forest fire early detection and segmentation, enabling the prediction of wildfire spread and supporting firefighting efforts.

The efficacy of standard computer vision approaches for wildfire identification is greatly hampered by atmospheric disturbances, such as haze, smoke, and cloud cover [25], [25], [26], [27]. These environmental conditions result in lower image clarity during wildfire recording, hindering the proper separation of fire boundaries. Haze, smoke, and cloud cover provide variable degrees of optical interference, hiding key characteristics such as the fire's leading edge, smoke plumes, and regions of heightened thermal activity. Consequently, the existence of these atmospheric artifacts hinders the segmentation and classification of wildfire-related pictures, leaving typical computer vision algorithms ineffective for detecting fire-affected areas amid background clutter.

In this research, we apply DCP approach to solve the issues provided by haze and fog in static images. The DCP approach requires the estimation of a depth map, known as the dark channel, generated from ambient light. This depth map serves to characterize the scattering of light generated by air particles, delivering useful insights regarding the level of haze present in the scene. In addition, we compute a transmission map, where pixel values indicate the relationship between the seen picture and the underlying scene features, therefore designating regions less influenced by fog. Finally, using the ambient light and the transmission map, the algorithm estimates the scene radiance at each pixel.

To simplify simultaneous classification and segmentation tasks inside dehazed wildfire images, we present our suggested Inv-Net, constructed using a multitask learning approach. The Inv-Net uses a modified INN as its backbone in the UNet model, boosting its adaptability and effectiveness in tackling complicated image-processing tasks. One key challenge lies in developing an operation that is both location-specific and channel-agnostic, crucial for tasks such as image classification and segmentation. This requires a solution that captures spatial information while being invariant to the specific channels within the image. To address this challenge, we incorporate the INN layer into the Inv-Net architecture, which produces each kernel conditioned on specific spatial positions. This approach enables the seamless processing of variable-resolution input tensors. This addition secures the Inv-Net's potential for handling different input data characteristics, thus boosting its practical usefulness in real-world scenarios. Unlike typical convolutional layers, INN layers dynamically alter their receptive fields based on input data properties, boosting the flexibility of feature processing. By jointly training both tasks inside a cohesive framework, Inv-Net capitalizes on the synergies between classification and segmentation, exploiting shared feature maps to boost the accuracy and robustness of both tasks. Through collaborative learning throughout the training phase, the classification and segmentation branches repeatedly refine each other's predictions, resulting in an enhanced overall performance.

Our main contributions can be summarized as follows.

1) Our research explores smart strategy by integrating the DCP method with wildfire data analysis. This fusion improves haze removal, enabling notable clarity in visualizing wildfires. The DCP algorithm achieves the highest average SSIM of 0.7831 and PSNR of 17.4321 demonstrating superior image quality compared to the other algorithms.
2) In addition, our technology, which incorporates segmentation and classification tasks into the Inv-Net architecture, improves wildfire recognition and segmentation accuracy. We enhance analysis for a more comprehensive understanding of wildfire image identification by using better feature extraction and involutional layers in the Inv-Net architecture. Inv-Net succeeds at both tasks, with 98.1% classification accuracy and an impressive 97.6% IoU score in segmentation, particularly in wildfire image processing.
3) Our study represents an advancement over previous computer vision techniques, permitting swifter and more precise wildfire identification and categorization. When benchmarked against other state-of-the-art deep learning algorithms, our methodology regularly yields satisfactory results, highlighting its usefulness and dependability in wildfire burn image analysis.

The following is a summary of the upcoming sections of this article. In Section II, we compare our findings to previous studies, emphasizing the differences and contributions of our proposed technique. Section III describes our collection of data approach, while Section IV defines DCP's preprocessing networks and provides the results gained when compared to other techniques. Section V describes our proposed multitasking approach, including the overall system architecture and functionality. Section VI covers our experimental setup, followed by Section VII, which discusses our model's performance measurements. In Section VIII, we examine our findings and insights to determine the effectiveness of our models. Finally, Section IX concludes by offering the derived results from our analysis and identifying potential directions for future research.

II. Related Work

This section focuses on recent advances in wildfire detection using deep learning methods. We discuss various traditional techniques employed for wildfire classification and segmentation tasks.

Kalaivani et al. [28] proposed a method that leverages transfer learning from MobileNet-V2 for forest fire detection. The Antlion Optimization (ALO) objective function is embedded within the parametric ReLU of the CNN layer. This design aimed to enhance accuracy and reduce error rates. The study used Landsat satellite images from five different forest fire regions and the custom-optimized CNN successfully detected various forest fire images. Azami et al. [29] used the MiniVGGNet network, which was employed for image classification. On-ground testing demonstrates that the proposed system achieved an overall accuracy of 98% and an F1 score of 97% in classifying wildfire events. The LeNet and ShallowNet networks were developed together and compared the classification results on wildfire images. Rashkovetsky et al. [30] proposed a workflow using deep learning semantic segmentation models. The research investigated the resources by combining data from different satellite sources: Sentinel-1, Sentinel-2, Sentinel-3, and MODIS. A single-input convolutional neural network (CNN) based on the U-Net architecture is trained for each satellite source.

Kang et al. [32] employed both random forest (RF) and (CNN) for model development in this research. This work contributes to forest fire detection by combining geostationary satellite data, reducing detection latency, and exploiting CNN-based models for accurate and timely alerts. TKhryashchev et al. [33] presented numerical experiments for a developed deep learning algorithm: U-ResNet34 for the segmentation of images of wildfire. The training and testing procedures were conducted using high-resolution aerial imagery sourced from the Resurs and Planet databases. According to test results for both datasets, U-ResNet34 worked satisfactorily. Abdusalomov et al. [34] introduced a modified forest fire detection model using Detectron2. The model showed high accuracy in fire detection scoring better performance. By exploiting the self-attention mechanism in transformers, the proposed model was able to effectively detect fire pixels with fewer parameters and reduced computational complexity.

Pan et al. [35] increasingly used deep learning for dynamic texture recognition, including wildfire detection. Inspired by previous work, this study proposed using MobileNet-V2 for forest fire detection. The integration of deep learning techniques with Fourier analysis is innovative in this research. By analyzing frequency responses, the model identified redundant or less informative filters. Zhang et al. [36] focused on comparing and analyzing the application of two feedforward neural network models (CNNs and MLPs) in global wildfire susceptibility prediction. Here, the contextual-based CNN-2D model performed best, utilizing neighborhood information effectively. Hongy et al. [37] proposed an addition-based efficient neural network for forest fire detection over a traditional CNN model. This AddNet performed 12.4% faster than CNN because AddNet did not use expensive matrix multiplication. Substantially it is almost binary-weight neural networks, but AddNet got lower true detection accuracy in this forest fire detection. Wang et al. [38] developed an optimized wildfire image classification framework based on Reduce-VGGnet and a wildfire region detection algorithm based on the optimized CNN with the combination of spatial and temporal features.

Jonnalagadda et al. [39] attempted to improve the detection of wildfires in real-time using drone/AAV pictures by utilizing a segmented neural network (SegNet) with a CNN architecture. This involved segmenting high-resolution images, implementing L2 regularization, and adopting data augmentation techniques. Ghali et al. [40] introduced CT-Fire, a novel ensemble approach that combines RegNetY with EfficientFormer v2. It accurately identifies tiny wildfires in the presence of complicated backgrounds and image quality limitations. Prakash et al. [41] developed an advanced early warning system for identifying

TABLE I
SUMMARY OF WILDFIRE DETECTION STUDIES

Ref	Objective	Method	Results	Limitations	MultiTasking	DCP Apply
[28]	Forest fire detection	Custom CNN, PReLU	Val. acc: +50-55%; Val. loss: 0.066; Train acc: 100%	limited dataset with an imbalanced class distribution	✗	✗
[29]	Wildfire classification	MiniVGGNet	Acc: 98%; F1 score: 97%	Needs fusion with image segmentation	✗	✗
[32]	Satellite image segmentation	Semantic segmentation	Sentinel-2 F1: 0.83; Combined models: +25% precision, +9% detection	sensor bias with no adaptive learning method	✗	✗
[34]	Forest fire detection	Dark channel prior, Detectron2	Prec: 99.4; Recall: 99.5; Acc: 99.4	Ineffective for similar atmospheric light scenes	✗	✗
[35]	Wildfire detection	MobileNet-V2	Det. rate: 91.60%; FA rate: 4.67%; Acc: 93.36%	Comparison with non-open source methods along with image resizing issues	✗	✗
[36]	Global wildfire susceptibility	CNNs, MLPs	CNN-2D :precision 96.76, Recall 94.60, F1 Score 95.67, Accuracy 95.71	Ignition data uncertainty with no clear ground truth	✗	✗
[37]	Wildfire detection	AddNet, window-based	Det. rate: 91.60%; FA rate: 4.67%; Acc: 93.36%	Comparison difficulties with unspecified training settings	✗	✗
[38]	Wildfire classification	Optimized CNN, Reduce-VGGNet	Class. acc: 91.20%; Region det. acc: 97.35%	Labor-intensive annotation with resizing affects accuracy	✗	✗
[39]	Wildfire detection	CNN architecture	Class. acc:98.18%; Recall: 98.74%	Potential loss of contextual information when segmenting images	✗	✗
[40]	Wildfire classification	deep CNN and the vision transformer	Class. acc: 99.62%	Computationally Expensive	✗	✗
[41]	Wildfire detection	Inception-v3 transfer learning model	Class. acc:97.55%; Recall: 96.44%	Model may not perform well with limited or imbalanced data	✗	✗
[42]	Wildfire Detection	MobileNet Transfer Learning Model	Class. acc: 70.00%	Network prediction performance did not noticeably improve changing different parameters tuning	✗	✗
[43]	Wildfire Detection	Convolutional Neural Network (CNN) with attention mechanisms	Class. acc: 95.51%	Model did not perform well on large dataset	✗	✗
Our Method	Wildfire Classification and Segmentation	Proposed Inv-Net Architecture	Classification acc: 98.1% and Segmentation acc: 97.6%	Overcomes the performance by integration of adaptive learning approach with different environmental situations	✓	✓

forest fires by employing a deep learning technique that utilizes a convolutional neural network (CNN) and transfer learning. The applied Inception-v3 model included preprocessing a dataset of 950 fire and 950 no-fire images. James et al. [42] examined the approach of applying the MobileNetV2 transfer learning model for wildfire identification in satellite imagery. They emphasized effective CNN optimization and hyperparameter adjustment for increased computational performance. Sifat et al. [43] developed CNNs with attention mechanisms for wildfire identification using satellite imagery, attaining 95.51% accuracy.

Existing research on wildfires reveals a major gap in the field of classification and segmentation algorithms. Also, some studies lack the diverse condition analysis in their model development process While most research focuses on discrete activities within this domain, there is an absence of emphasis on dynamic techniques that cater to the evolving nature of wildfires. In addition, many investigations miss the critical aspect of haze removal, particularly for aerial images where haziness in the sky can greatly hamper the accuracy of classification and segmentation models. Table I summarizes the existing studies and limitations. Standard computer vision systems have made considerable progress in detecting wildfires, but they are ineffective in real-world situations where environmental circumstances can rapidly change. Standard CNNs and simple picture segmentation methods frequently encounter difficulties when dealing with diverse spatial characteristics and atmospheric disturbances such as haze, smoke, and cloud cover. Fixed CNN models may demonstrate efficacy in controlled environments but lack the adaptability required for real-time wildfire monitoring [44], [45]. Multiple studies have emphasized these constraints, pointing out that stationary models are insufficient in handling wildfires' ever-changing and uncertain characteristics.

The adaptive multitask learning framework is a fundamental element of our method. By merging classification and segmentation tasks into a single deep-learning model (Inv-Net), we enable the network to learn from both tasks simultaneously. This combined learning process allows the model to constantly alter

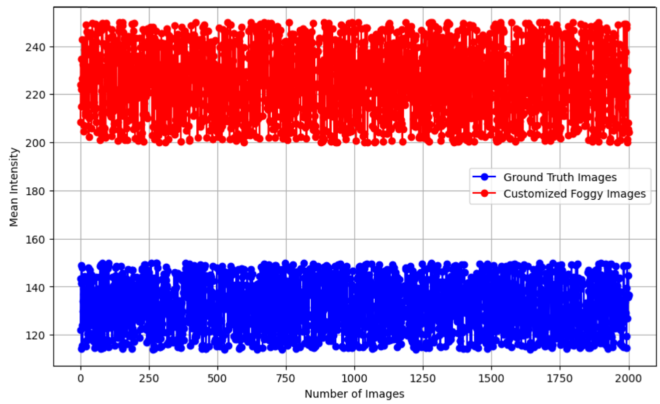

Fig. 2. Comparing the average intensity of original images with customized hazy versions.

its predictions based on new information, continually improving its accuracy for both tasks. The shared feature maps built during training promote higher generalization and flexibility. Also, it makes the model more responsive to changes in wildfire behavior involutional layers, unlike fixed convolutional filters, actively modifying their receptive fields according to input data attributes. This flexibility provides well under diverse settings such as different resolutions, fire intensities, and environmental interferences.

III. Dataset

Our study involves a two-stage approach. Initially, a preliminary stage is devoted to data preprocessing, utilizing the DCP technique. The second stage of our proposed multitask model focuses on the categorization and segmentation of wildfire aerial images. Fig. 12 depicts the overall process utilized in our study.

Thus, our primary focus is on the process of collecting data. For the data collection process, for our classification and segmentation task for our task, we introduce the fire luminosity airborne-based machine learning evaluation (FLAME) dataset: a rigorously curated collection of aerial pictures and videos collected via AAVs during controlled pile burn operations within a Northern Arizona pine forest [31]. Tailored to support breakthroughs in fire detection and management algorithms, this dataset contains high-resolution video footage of pile burns fully merged with associated thermal pictures, thus catching key temperature variations. Each frame in the dataset is properly labeled to show the presence or absence of fire, giving vital help for algorithmic training initiatives. A total of 39 375 images have been identified as either “Fire” or “Non-Fire” during the training phase. In addition, 8617 images had been labeled for the test data. A total of 2003 images were analyzed for fire segmentation. As a result, 2003 masks were created to provide ground truth data with pixelwise annotation. The image sizes are generated as 3840×2160 pixel value. Once the dataset is acquired, each image is loaded using the OpenCV library. This procedure guarantees that the images are available for additional analysis and alteration. In the situation that an image is not located at the specified path, adequate error-handling mechanisms are in place to handle such circumstances gracefully, ensuring the smooth execution of the prepossessing pipeline. Afterward, from the collected framed static data, we focus on the robustness of your proposed model by applying the haze addition process.

A. *Fog Addition*

To replicate a foggy appearance in images, we utilize the approach put forward by Godard et al. [46]. This approach involves the estimation of object depth using only one image, which is achieved by self-supervised training. During this training, the model acquires the ability to forecast perspective by using another image. It utilizes the difference between two images to generate a depth map (D_t). Subsequently, we employ (1) to measure the magnitude of the error and enhance both the ultimate image and its production procedure [46]

$$L_p = \min_{p} e(I_t, I_{t'} \rightarrow t) \cdot t' \tag{1}$$

where L_p indicates the error in photometric projection and prephotometric reconstruction, estimated using SSIM and L1 distance. SSIM and L1 distance form components of the photometric restoration error, determined from multiple research sources [47], [48], [49]

$$e(I_a, I_b) = \frac{\alpha}{2}(1 - \text{SSIM}(I_a, I_b)) \quad + (1-\alpha)||I_a - I_b|| \tag{2}$$

where $\alpha = 0.85$ reflects the weight of the wall. In addition, we use edge-aware smoothness to achieve smooth depth maps, as represented in

$$L_s = |\partial_x d_t| e^{-|\partial_x I_t|} + |\partial_y d_t| e^{-|\partial_y I_t|}. \tag{3}$$

Furthermore, by Godard et al. [46], we apply some approaches, including prepixel minimum re-projection, to increase the precision of projected depth maps without increasing model complexity. To tackle scenarios with moving objects, we present two strategies: minimum error and automasking stationary. Here, (4) dynamically calculates a binary mask μ by taking into account projection errors and multiscale estimation, which helps avoid the optimization process from being stuck in local minima

$$\mu = \left[\min_{t'} e\left(I_t, I_{t'} \rightarrow t\right) < \min_{t'} e\left(I_t, I_{t'}\right)\right]. \tag{4}$$

After deriving depth maps from the original images and inverting them, we configure the model to import the designated target image for simulating foggy conditions. Furthermore, in Fig. 4, the depth maps of images are shown in a 3-D construction map for a better understanding of the inconsistencies of the pixels. In a standard image without haze, the pixel values span from 0 to 255, representing the level of clarity in the color of each pixel. Images that have a moderate level of clarity often show noticeable differences between bright and dark areas. In foggy images, brighter pixel values dominate while darker values diminish due to reduced visibility induced by the fog. In Fig. 2, an obvious finding is the discrepancy in mean intensity values between the ground truth image and the monocular depth-based hazy image. This mismatch gives solid evidence of true fog existence within the modified sample.

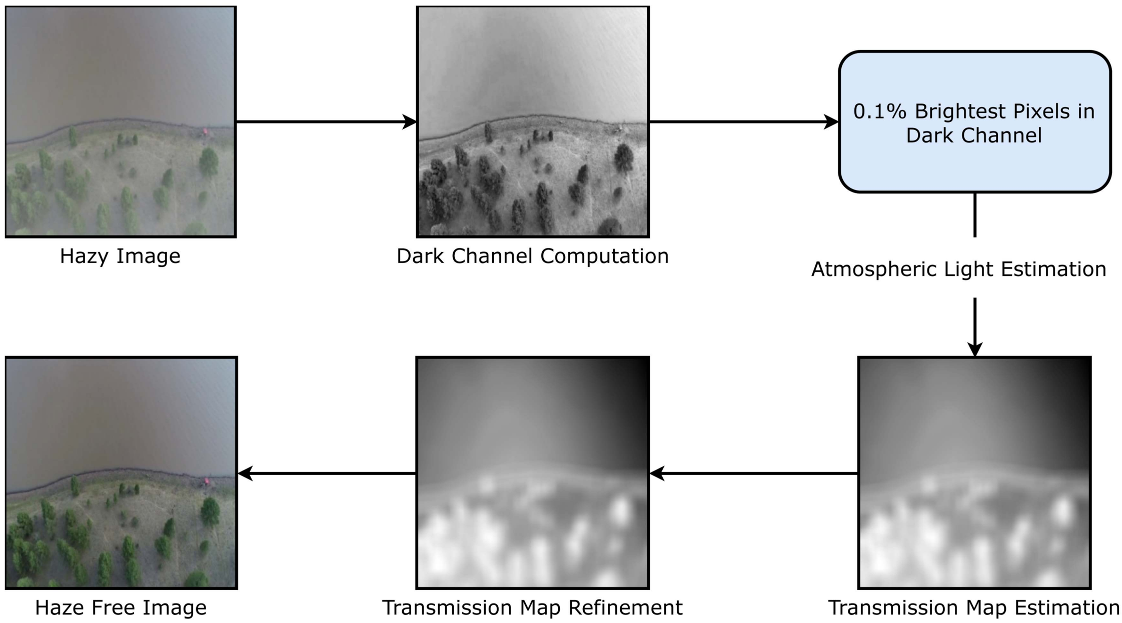

Fig. 3. Implementing DCP-based algorithm for defogging of our Wildfire Images.

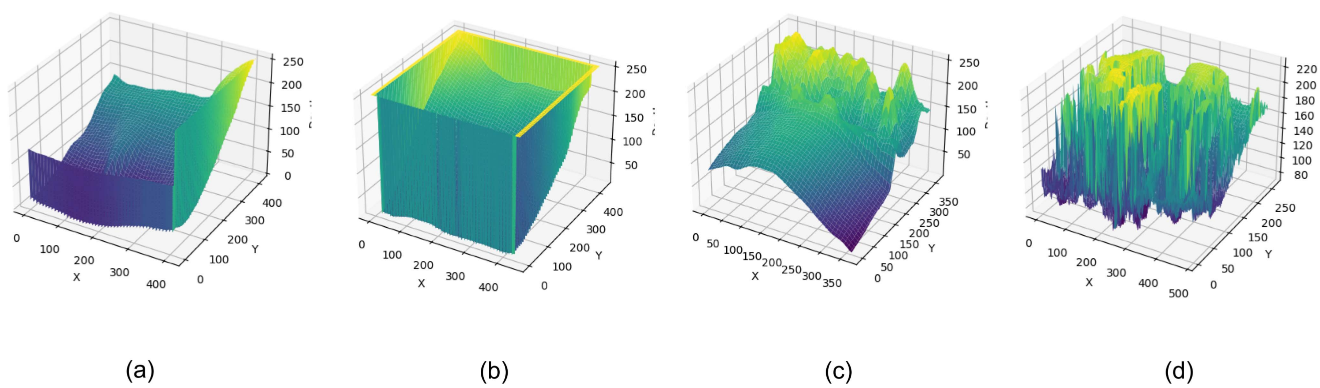

Fig. 4. 3D map construction of depth Images for samples captured at different locations: (a) location-1; (b) location-2; (c) location-3; (d) location-4 .

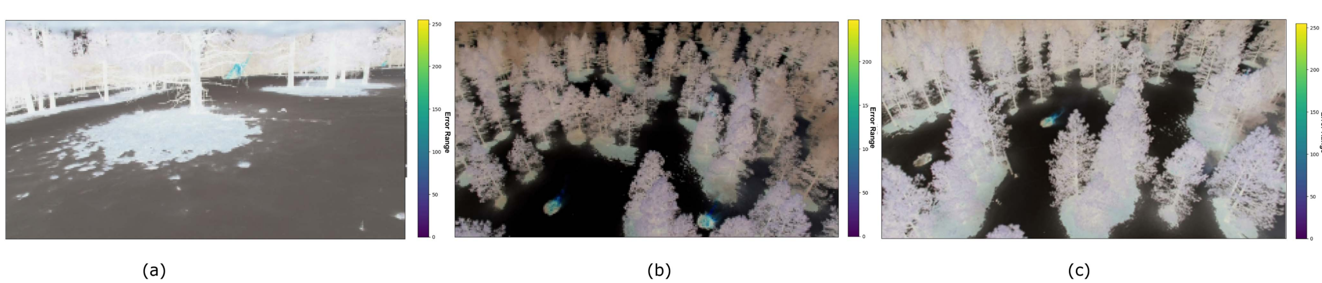

Fig. 5. Error map construction of sample dehazed images: (a) location-1; (b) location-2; (c) location-3 .

B. Smoke Addition

To include smoke with different spatial distributions in the dataset, we blend a smoke-colored layer [with RGB values ranging from light smoke at (200, 200, 200) to dense smoke at (100, 100, 100)] with intensities ranging from 0.3 to 0.7, optionally with Gaussian noise. For capturing realistic smoke effects, this combination is modified with a vertical gradient and an optional fire mask as shown in Fig. 7 labeled as 1) light smoke, 2) Medium Smoke, 3) Dense Smoke, 4) Colored Smoke. In differentiating

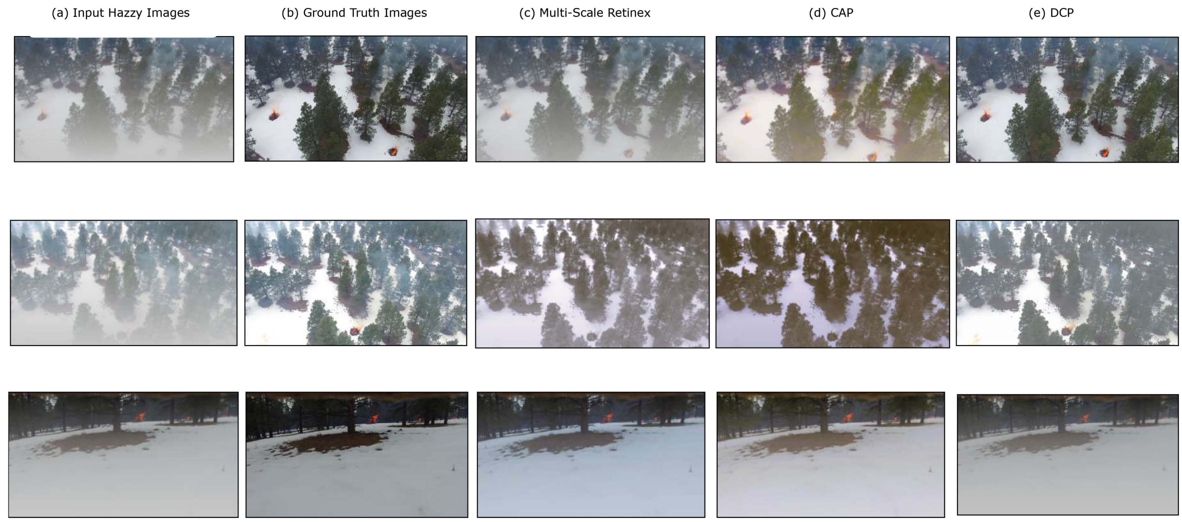

Fig. 6. Comparative analysis of dehazing algorithms, (a) original hazy images, (b) the corresponding ground truth images, (c) multiscale Retinex images, (d) CAP images, (e) images obtained from the DCP algorithm.

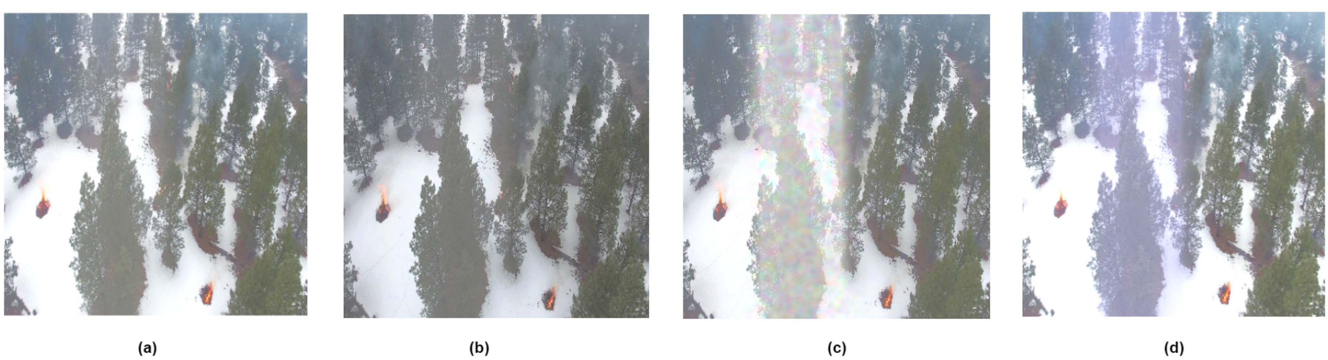

Fig. 7. Integrating smoke layers with different RGB values and intensities to create authentic smoke effects. The variations consist of (a) light smoke, (b) medium smoke, (c) dense smoke, and (d) colored smoke.

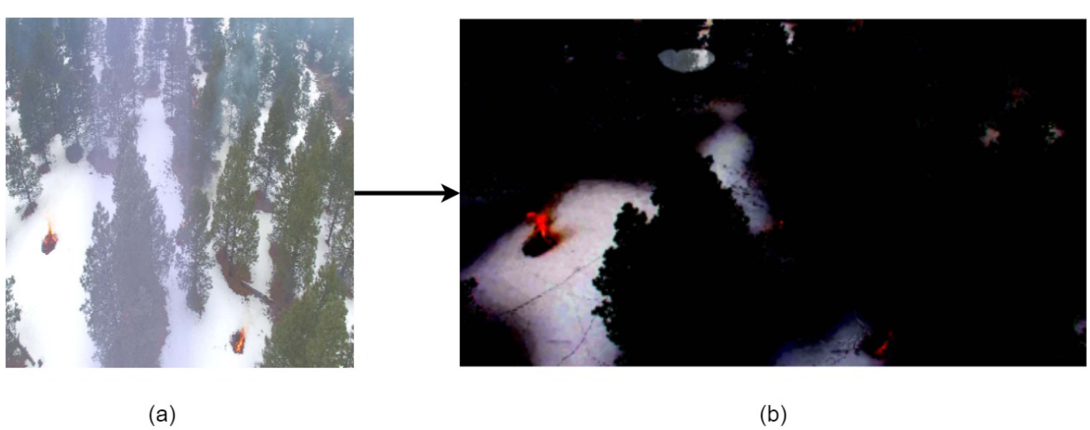

Fig. 8. Utilizing DCP for targeted preservation: (a) Original hazy images for retention; (b) Specific focus on smoke-affected regions.

between haze and smoke, we further apply the DCP approach to estimate the atmospheric light and transmission map. We then utilize the threshold mask to retain smoke and fire regions by applying the mask to the original image. This procedure ensures that haze is properly removed while maintaining crucial visual characteristics like smoke and fire in Fig. 8. After applying the DCP dehazing process, a similar approach is used for smoke removal.

IV. Data Preprocessing Using DCP

In this study, we propose the utilization of the DCP technique for image-defogging purposes. The DCP approach comprises the estimation of a transmission map, which serves as a vital component in the restoration of haze-free images. This approach operates on the idea that local regions within hazy photos containing low-intensity pixels throughout at least one color channel are suggestive of haze-free locations. Leveraging this assumption, the dark channel is generated from the hazy image, outlining the least intensity values across all color channels within localized image regions. Subsequently, the transmission map is obtained from the dark channel, simplifying the elimination of haze by guiding the application of an atmospheric scattering model. Fig. 3 displays the process of applying the DCP method for the haze removal process. The method demonstrates a substantial reduction in haze, hence producing perceptually crisper representations.

A. Patch Selection

In this phase, the algorithm begins by picking a local patch $\omega(x)$ centered at each pixel x in the input low-light image I. This patch acts as a local region of interest for further investigation.

B. Dark Channel Calculation

Once the local patch is picked, the dark channel $J_{\text{dark}}(x)$ at pixel x is computed. This is performed by determining the minimal intensity value within the patch $\omega(x)$ shown in (5). Mathematically, it is expressed as

$$J_{\text{dark}}(x) = \min_{y\in\omega(x)} (I(y)) \tag{5}$$

where y iterates over all pixels in the patch $\omega(x)$. This step effectively identifies the darkest regions within the local patch, providing valuable information about the overall darkness in the image. Here in our DCP, we generate a function that is written to calculate the dark channel of each image in the collection. This function computes the minimum intensity value across all color channels (red, green, and blue) at each pixel location, resulting in a dark channel image that accentuates the presence of haze or smoke in the original image.

C. Atmospheric Light Estimation

Following the calculation of the dark channel, our DCP method proceeds to estimate the ambient light A shown in (6). This value is vital for accurately representing the scene's illumination. The ambient light is determined as the pixel with the highest intensity in the dark channel:

$$A = \arg\max_{x} (J_{\text{dark}}(x)) . \tag{6}$$

By selecting the pixel with the highest intensity in the dark channel, the algorithm efficiently captures the brightest spot in the picture, which often correlates with the ambient light. In our DCP method, the ambient light is based on the dark channel of each image. This procedure comprises selecting the top 0.1% brightest pixels from the dark channel and determining the greatest intensity value among these pixels. Algorithm 1 shows the process.

Algorithm 1: Atmospheric Light Estimation.

Input: I: Input low-light image, D: Dark Channel Prior Image
Output: A: Estimated Atmospheric Light
Compute R, C (dimensions of I).
Initialize $M = 0$.
for $r = 1$ **to** R **do**
  for $c = 1$ **to** C **do**
    if $D(r, c)$ is max in its neighborhood and $M < I(r, c)$ **then**
      Set $M = I(r, c)$ and $A = I(r, c)$.
    end if
      end for
end for
return A.

D. Transmission Map Calculation

The transmission map t is computed by leveraging both the atmospheric light channel and the dark channel shown in (7). Here, $\tilde{t}(\mathbf{x})$ is defined as

$$\tilde{t}(\mathbf{x}) = 1 - \min_{i} \left(\min_{\mathbf{p}\in\Omega(\mathbf{x})} \left(\frac{I_i(\mathbf{p})}{A_i} \right) \right) \tag{7}$$

where I_i represents the normalized input image divided by the atmospheric light A_i, and Ω (omega) is a constant. Essentially, $\min_i(\min_{\mathbf{p}\in\Omega(\mathbf{x})} I_i(\mathbf{p})/A_i)$ denotes the dark channel of the normalized haze image $I_i(\mathbf{p})/A_i$, providing an estimation of the transmission. The transmission map provides insights into the visibility conditions within the image and plays a critical role in scene radiance recovery. A function is designed to estimate the transmission map using the normalized picture and the ambient light. This entails calculating the raw transmission based on the dark channel of the normalized image and applying a blur to smooth the transmission map. The omega parameter controls the influence of the atmospheric light on the transmission map, allowing for fine-tuning of the recovery process. Besides, Fig. 5 depicts the error map and illustrates the gap between the original image and the image after haze reduction. Brighter sections suggest more disparities, revealing locations where the haze removal technique may have generated artifacts or inaccuracies. By comparing the error map with the original image, we spot places needing improvement and tweak the haze removal algorithm accordingly.

E. Transmission Map Refinement

Refining the transmission map through guided filtration, as proposed by He et al. [50], involves using guided filtering to enhance edges and preserve essential details. The guided filter takes the grayscale image and the estimated transmission map as inputs to generate a refined transmission map $t(\mathbf{x})$. $t(\mathbf{x})$ represents the refined transmission map shown in (8). To minimize the cost function, we express $t(\mathbf{x})$ and $\tilde{t}(\mathbf{x})$ as t and $\tilde{t}$, respectively,

TABLE II
DEHAZING ALGORITHMS PERFORMANCE COMPARISON

Algorithm	Average SSIM	Average PSNR	Execution Time (s)
DCP	0.7831	17.4321	0.4257
Multi-Scale Retinex	0.6412	15.4234	10.5618
CAP	0.5324	14.879	2.327

in vector form:

$$E(\mathbf{t}) = \mathbf{t}^T L\mathbf{t} + \lambda \left(\mathbf{t} - \tilde{\mathbf{t}}\right)^T \left(\mathbf{t} - \tilde{\mathbf{t}}\right) \tag{8}$$

where L is the Matting Laplacian matrix proposed by Levin, and λ is a regularization parameter.

F. Scene Radiance Estimation

Finally, using the ambient light A and the transmission map $t(x)$ mentioned in (9), the algorithm estimates the scene radiance $J(x)$ at each pixel x. This stage involves adjusting the original pixel intensities based on the transmission map and atmospheric light:

$$J(x) = \frac{I(x) - A}{\max\left(t(x), t_0\right)} + A \tag{9}$$

where t_0 is a small constant introduced to prevent division by zero and ensure numerical stability. By applying this equation, the algorithm enhances the visibility and clarity of the low-light image, revealing previously obscured details.

G. DCP Assessment and Comparison

Our experimental setting comprises the implementation of two different models: Multi-Scale Retinex [51], and Color Attenuation Prior (CAP) [52] model to evaluate their efficacy with the DCP method. Table II provides the performance characteristics of three different dehazing algorithms: Dark Channel Prior DCP, and multiscale retinex (CAP). The metrics include the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR), which are measures of image quality, as well as the execution time of each algorithm. In Fig. 6, a comparative comparison of dehazing methods presents original hazy pictures, related ground truth, Multi-Scale Retinex enhanced images, CAP-dehazed images, and images processed using the DCP algorithm. From the data, it is obvious that the DCP algorithm obtains the highest average SSIM of 0.7831 and PSNR of 17.4321, showing greater image quality compared to the other algorithms. In addition, DCP displays the fastest execution time among the algorithms, with an average of 0.4257 s. On the other hand, multi-scale retinex exhibits lower SSIM and PSNR values, implying comparably weaker dehazing performance. CAP also shows lower SSIM and PSNR scores, where it shows with a moderate execution time. These findings illustrate the usefulness of the DCP method in generating both high-quality dehazing outcomes and efficient computing performance. For assessing the ablation performance, in Fig. 9, we analyze the sensitivity of the omega variable in the atmospheric light estimation of the DCP technique. Our observations reveal that at an omega value of 0.95, the approach exhibits excellent performance, reaching its highest effectiveness.

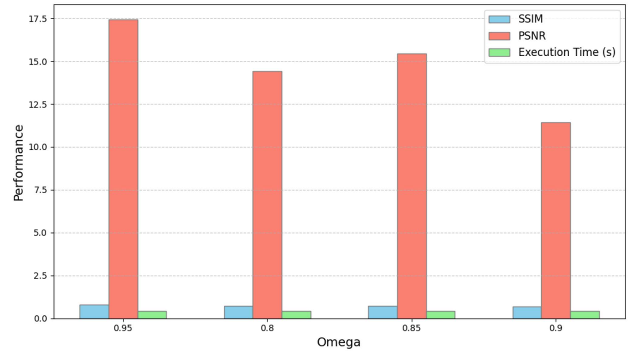

Fig. 9. Impact of omega on the performance of light transmission estimation.

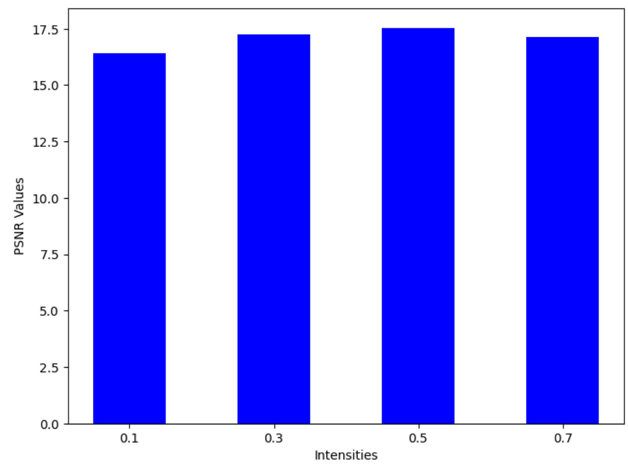

Fig. 10. Impact of different smoke color intensities on the performance of DCP model dehazing smoke haze process.

V. METHODOLOGY

In this study, we present a novel approach to wildfire segmentation and classification tasks using the DCP method and our multitasking Inv-Net model. The dehazed image data undergoes initial processing before being sent into the Inv-Net intended for wildfire segmentation and classification. The architecture of the Inv-Net builds on the U-Net framework and is structured into distinct segments: the encoding path, the classification head, the decoding path, and the segmentation head. These components collectively contribute to the comprehensive analysis and categorization of wildfire-related imagery. Here in Fig. 11 displays the whole proposed workflow utilized in our study. Furthermore, in Fig. 12 depicts our proposed Inv-Net model used in our study.

A. Encoding Path

The encoding element of the architecture employs downsampling blocks incorporating INN layers, batch normalization, and Leaky ReLU activation functions. These blocks collect high-level features by progressively lowering the spatial dimensions of the input feature maps while increasing their depth. INN layers are exploited for their ability to efficiently capture global context with fewer parameters compared to classic convolutional layers. Each downsampling block consists of an INN layer followed by batch normalization and leaky ReLU activation,

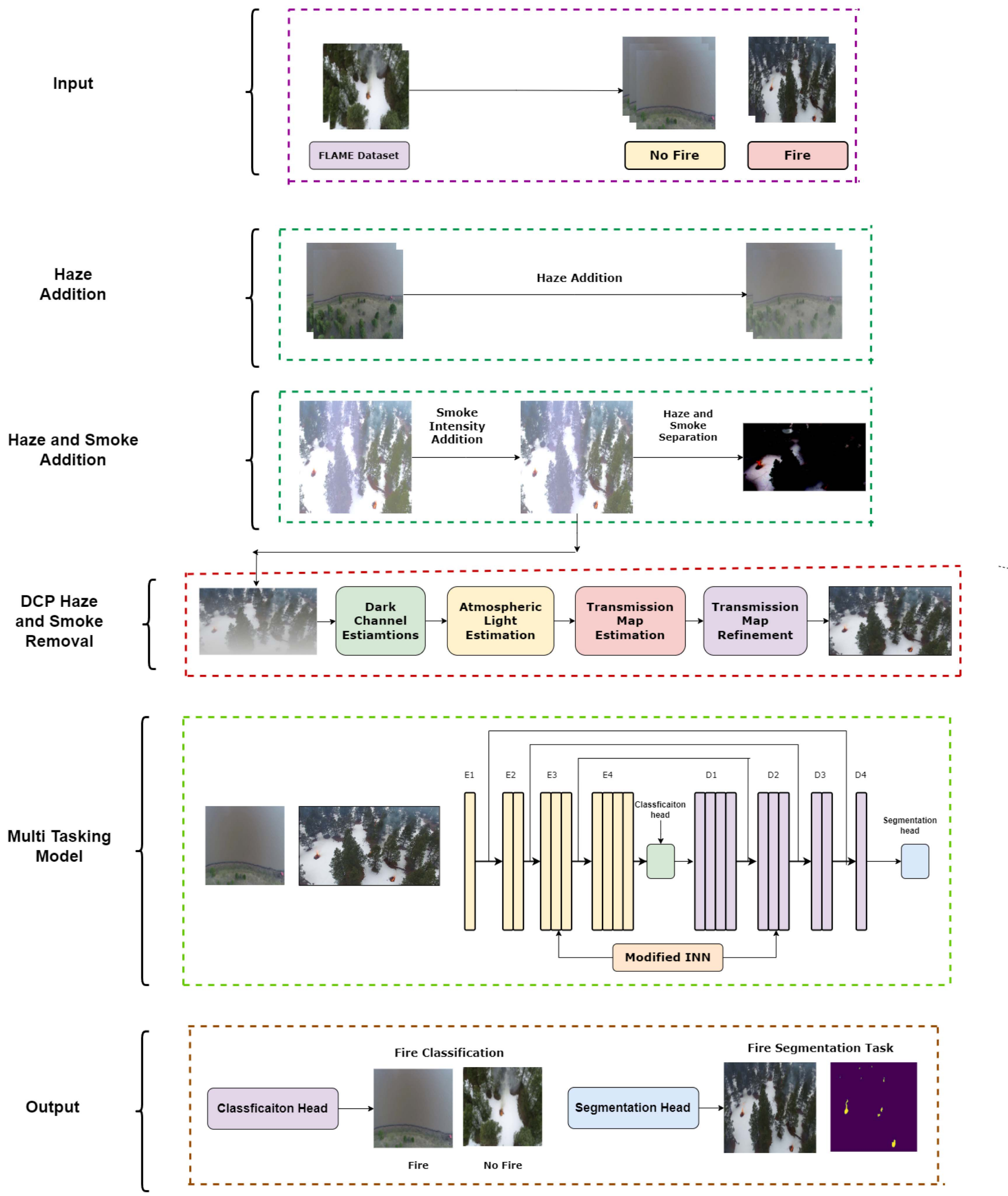

Fig. 11. Proposed workflow.

allowing the extraction of discriminative features. The output feature maps from the encoding path are later shared with the classification and segmentation heads for additional processing. This hierarchical feature extraction technique enables the model to efficiently gather and represent crucial information necessary for wildfire categorization. The feature map count starts at 3, doubling with each downsampling step, and is limited to 64 to manage computational resources.

B. Classification Head

The classification head, positioned at the base layer of the architectural framework, is defined by a convolutional block succeeded by global average pooling, adeptly catering to inputs of varying dimensions. Subsequent layers comprise fully connected layers, reinforced with leaky rectified linear unit (ReLU) activation, culminating in a softmax layer, thereby producing evaluation probabilities relevant to fire classification. In this classification task, the INN Layer successfully integrates global context, consequently increasing comprehension of the encompassing setting, a vital component for identifying diverse levels of input images. Moreover, its parameter efficiency stands out, since it performs global context aggregation with fewer parameters compared to typical convolutional layers [53]. This not only boosts the model's performance but also mitigates the risk of overfitting, assuring solid predictions during both the training and testing phases.

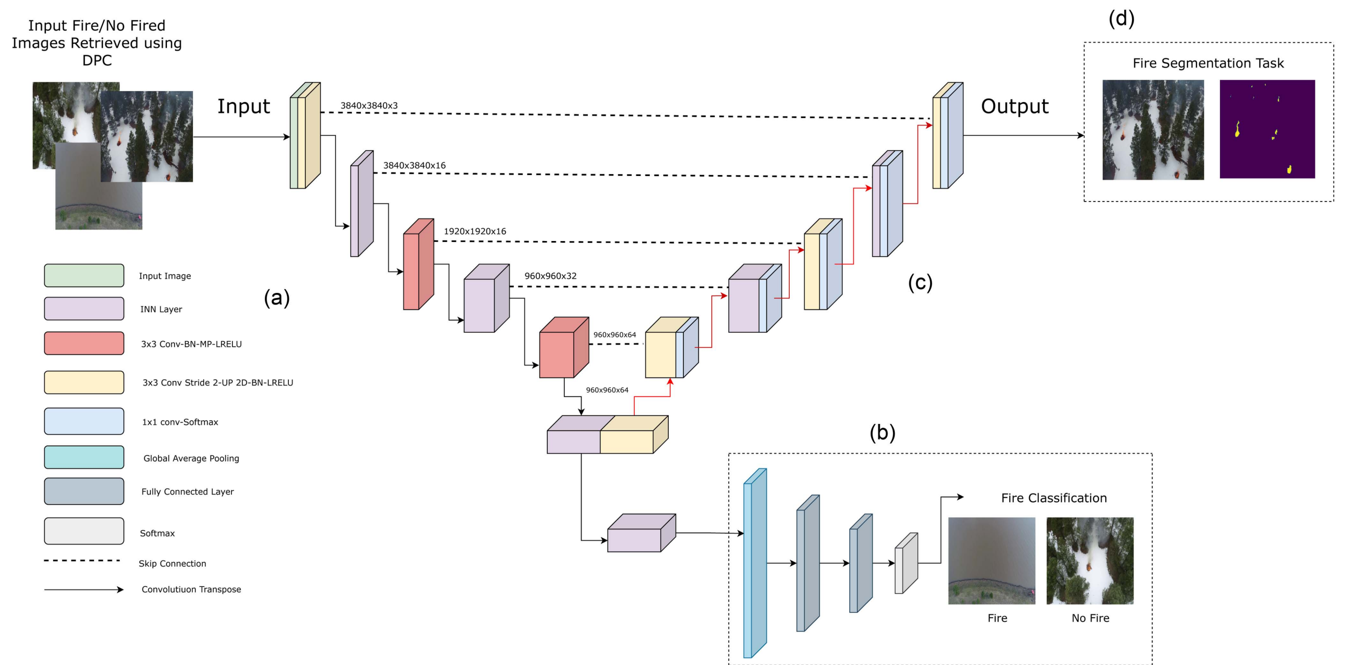

Fig. 12. Illustration depicting the components of the proposed Inv-Net for the classification of wildfires and segmentation of lesions. The diagram showcases (a) the encoding path, (b) the classification head, (c) the decoding path, and (d) the segmentation head.

C. Decoding Path

The decoding pipeline of the Inv-Net is characterized by the utilization of upsampling blocks integrating two INN layers, which are crucial in reinstating feature maps to their original input dimensions, thereby facilitating the elucidation of low-level data attributes. Each block is characterized by the addition of two INN layers alongside batch normalization and leaky (ReLU activation. Intervening between blocks, subsampling is executed by transposed convolutions, ensuring a harmonized connection between the encoding and decoding pathways. To sustain the accuracy of spatial information, skip connections are applied, whereas feature maps from encoding and decoding blocks are combined. The upsampling blocks comprising INN layers serve the primary goal of restoring feature maps to their original input dimensions.

D. Segmentation Head

In the segmentation head of Inv-Net, feature mappings emanating from the final decoding block are processed through two consecutive INN layers. Each of these levels is separately designed with channel configurations of 32 and 16, respectively. The choice of applying a 3×3 kernel size while maintaining a stride value of 1 improves the effective capture of information from a broader environmental context. Subsequently, batch normalization is employed to normalize activations, hence, assuring stable training circumstances. Concurrently, the employment of Leaky ReLU activation functions brings nonlinearity into the network, hence, facilitating effective learning processes. Finally, a softmax layer is applied to generate a probability distribution across classes, hence, ensuring the interpretability of segmentation predictions and facilitating precise segmentation outputs. The INN layer in the segmentation head is highly skilled at capturing complex spatial connections. In addition, the INN layer's involvement in deep supervision improves training efficiency by enabling the construction of probability maps at each decoding block. This optimization of computational resources is achieved through parameter efficiency [53].

E. Loss Function

We employ a dual-task (10) approach, combining classification and segmentation components [54], [55]. For classification, we utilize multiclass cross-entropy loss, defined as

$$L_{\text{cls}} = -\sum_{k=1}^{K} y_k \times \log\left(\hat{y}_k\right) \tag{10}$$

where y represents ground truth labels in one-hot format for each class k, and $\hat{y}_k$ is the softmax output of the classification head for class k.

In segmentation, we integrate cross-entropy (11) loss with the DICE loss to handle imbalanced data scenarios. The DICE loss [56], is defined as

$$L_{\text{DICE}} = -\frac{1}{K}\sum_{k=1}^{K} \frac{2 \times \sum_{i=1}^{N} \hat{m}_i^k \times m_i^k}{\sum_{i=1}^{N} \hat{m}_i^k + \sum_{i=1}^{N} m_i^k + \delta} \tag{11}$$

focuses on the similarity between predicted and actual segmentation maps. Here, $\hat{m}_i^k$ represents the softmax output of the segmentation head for pixel i in class k, m_i^k denotes the one-hot encoded ground truth segmentation map for the same pixel, N is the total number of pixels per image, and δ is a small constant.

For deep supervision during segmentation training, the final segmentation loss shown in (12) where (L_{seg}) is a weighted sum of losses from all resolution outputs of Inv-Net. Specifically, it combines DICE loss and cross-entropy loss, denoted as L_{DICE} and L_{CE}, respectively. Thus, the segmentation loss is

$$L_{\text{seg}} = \sum_{l=1}^{L} w_{\text{seg}}^{l} \times L_{\text{seg}}^{l} \tag{12}$$

where w_{seg}^{l} represents the weight for the softmax output of the lth Inv-Net resolution, l ranges from 1 to L, and L signifies the total number of Inv-Net outputs.

To create a comprehensive loss function, we combine the classification and segmentation losses (13) with respective weighting coefficients α_{cls} and α_{seg}, yielding

$$L_{\text{Inv-Net}} = \alpha_{\text{cls}} \times L_{\text{cls}} + \alpha_{\text{seg}} \times L_{\text{seg}}. \tag{13}$$

This formulation ensures both tasks are adequately represented in the training process.

F. Post Processing

After obtaining the segmentation results from the Inv-Net segmentation head, the final predictions are obtained by choosing the label with the highest probability for each pixel. Following this, any solitary segmented pixels are deleted. Ultimately, the process culminates with the classification of fire and the segmentation of lesions.

VI. Experimental Setup

The assessment of the proposed Inv-Net entails conducting multiple experiments to evaluate its precision and robustness. The testing encompassed ablation research, comparison with conventional methods for classification, and segmentation tasks. In addition, comparisons were conducted with different multitask techniques. The training of all networks was conducted using the provided parameters, with the learning rate adjusted according to the training loss curve of each network. The experiments were performed using Python code on a hardware setup comprising an Intel i7 CPU operating at a frequency range of 2–8 GHz, accompanied by 16 GB of RAM. In addition, an NVIDIA GTX 3070 GPU with 16 GB of memory was utilized, making use of cuDNN 10.1 libraries. The implementation was carried out using the MONAI framework with the PyTorch library as the backend. The statistical significance of the classification findings was determined using a one-sided McNemar's test, while an overall accuracy comparison was conducted using a joint 2×2 table. For segmentation outcomes, statistical significance is tested versus other algorithms in terms of performance metrics using a one-sided paired t-test. The normality of the data was tested using the D'Agostino-Pearson test. For the multitask loss, α_{cls} and α_{seg} are set to 0.6 and 0.8, respectively, which showed to achieve better performance. The experimental setup of all the training parameters used in our study is shown in Table III.

TABLE III
Experimental Setup Summary

Component	**Specification**
Operating System	Windows 11
CPU	Intel i7 CPU
GPUs	NVIDIA GTX 3070 GPU
TensorFlow Version	2.11.0
Framework	PyTorch Monai
Total Training Epochs	100
Batch Size	32
Dropout Rate	0.4
Regularization	L1 and L2 kernel regularization techniques
Optimizer	Adaptive Gradient Algorithm (Adagrad)
Learning Rate	0.00001

VII. Performance Metrics

To evaluate the effectiveness of our proposed model along with other models, we employ widely used metrics: ACC (accuracy) (14), PRE (precision) (15), REC (16) (recall), F1-score (17), and IoU (18) (Jaccard index), Flop Count, Inference time. These metrics offer a comprehensive evaluation of model performance and facilitate comparisons with other studies in the field. All the formulas with definitions of the metrics are shown in Table IV

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$\text{PRE} = \frac{TP}{TP + FP} \tag{15}$$

$$\text{REC} = \frac{TP}{TP + FN} \tag{16}$$

$$\text{F1-score} = \frac{2 \times \text{PRE} \times \text{REC}}{\text{PRE} + \text{REC}} \tag{17}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN}. \tag{18}$$

Ablation studies are conducted to systematically evaluate the efficacy of individual components within the proposed method, specifically targeting optimization approaches, loss functions, deep supervision training, batch sizes, and the number of convolution kernels within the Inv-Net architecture. Initially, the dataset underwent random partitioning into training and testing subsets, comprising 80%, 20% of the whole dataset accordingly. Subsequently, each experiment entailed the methodical replacement of a single component while maintaining consistency in the remaining elements. The redesigned Inv-Net architecture is then subjected to training and evaluation on the corresponding training and testing sets. Furthermore, Inv-Net (1), Inv-Net (2), Inv-Net (3), and Inv-Net (4) models are conducted on various metrics for ablation studies shown in the block diagram in Fig. 13. We improve the performance of the INN layer through fine-tuning, utilizing larger kernels, and experimenting with compared strides. These modifications are crucial in achieving better outcomes. The parameters of all the modified versions of INN layers in the Inv-Net architecture are given in Table V.

TABLE IV
PERFORMANCE EVALUATION METRICS DEFINITIONS

Metric	Definition	Formula
Accuracy (ACC)	This metric measures the proportion of correctly classified instances out of the total instances in the dataset. It provides an overall indication of how well the model is performing.	$\mathrm{ACC} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$
Precision (PRE)	Measures the proportion of true positive predictions among all positive predictions made by the model.	$\mathrm{PRE} = \frac{\text{True Positives}}{\text{True Positives}+\text{False Positives}}$
Recall (REC)	Measures the proportion of true positive predictions among all actual positive instances in the dataset.	$\mathrm{REC} = \frac{\text{True Positives}}{\text{True Positives}+\text{False Negatives}}$
F1 Score	Harmonic mean of precision and recall.	$F1 = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$
Intersection over Union (IoU)	Calculates the ratio of the area of overlap between the predicted and ground truth regions to the area of the union of these regions.	$\mathrm{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$
FLOP Count	Represents the number of arithmetic operations performed by the model during inference or training.	–
Inference Time	It is the amount of time it takes a machine learning model to process input data and make predictions. It assesses the model's computational efficiency throughout the inference phase.	–

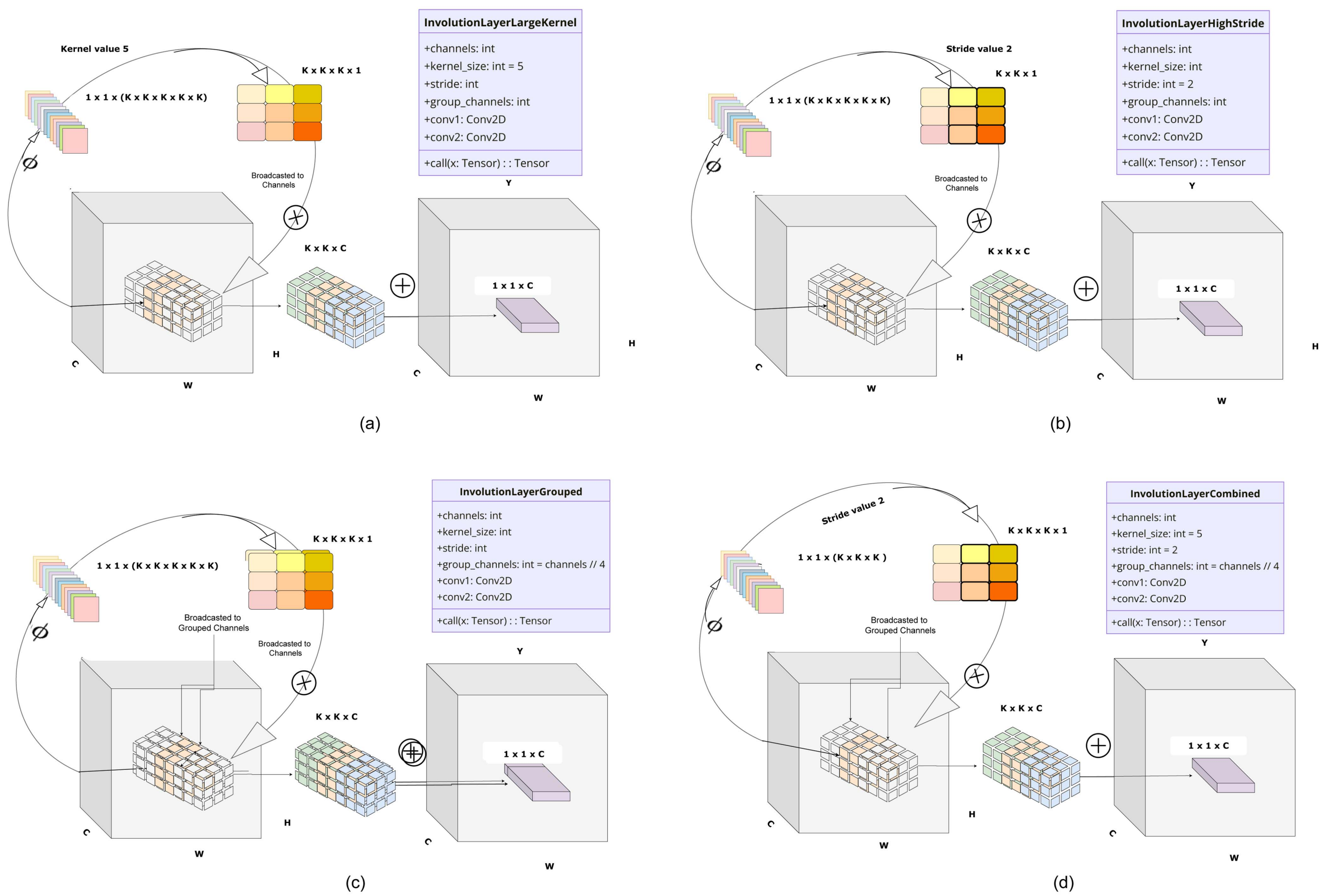

Fig. 13. Ablation studies of Inv-Net incorporating different INN layers applying different operations (a) Inv-Net (1) with higher kernel, (b) Inv-Net, (2) with higher stride, (c) Inv-Net, (3) with grouped channels, (d) Inv-Net, (4) with combined channels.

TABLE V
PARAMETERS OF INN LAYER MODELS

Model	Channels	Kernel Size	Stride
Inv-Net (1)	**8**	3×3	**1**
Inv-Net (2)	**16**	5×5	**2**
Inv-Net (3)	**32**	3×3	**1**
Inv-Net (4)	**64**	7×7	**2**

A. Classification Task

Within the field of classification problems, comparison analyses are presently being undertaken among important architectures such as ResNet101 [57], EfficientNetB4 [58], DenseNet161 [59], VGG16 [60], deep convolutional neural network (DCNN) [61], and the encoding block which used

TABLE VI
SUMMARY OF ABLATION STUDIES ON THE INV-NET MODEL, DETAILING THE MODIFICATIONS TO EACH KEY MODULE

Model	Segmentation				Classification			
	ACC	PRE	REC	IoU	ACC	PRE	REC	F1-score
Dropout 0.2	96.7	96.5	96.8	96.7	95.6	96.0	96.9	95.5
SGD	96.1	95.7	96.9	96.0	97.8	97.6	97.9	97.8
Batch size=128	95.2	94.7	96.0	95.0	97.1	96.8	97.4	97.1
Involutional Layer1 Freeze	96.5	96.2	96.7	96.3	95.9	95.8	94.5	95.9
Involutional Layer2 Freeze	96.8	96.4	96.6	96.8	95.2	94.9	95.5	95.2
Loss seg. only DICE	96.3	96.0	96.5	96.3	97.1	97.7	96.6	97.1
Loss seg. only CE	95.1	95.0	95.2	95.1	96.5	96.8	96.6	96.5
Proposed	98.2	97.9	97.5	97.1	97.9	97.7	98.3	97.9

TABLE VII
ABLATION STUDIES ON THE INV-NET MODEL'S PERFORMANCE, SHOWCASING ADJUSTMENTS TO INN MODELS WITH DIFFERENT HYPERPARAMETER TUNING, INCLUDING KERNELS, STRIDES, AND CHANNELS

Model	Segmentation				Classification			
	ACC	PRE	REC	IoU	ACC	PRE	REC	F1-score
Inv-Net (1)	95.1	95.0	95.2	95.1	96.5	96.8	96.6	96.5
Inv-Net (2)	96.1	95.7	96.9	96.0	97.8	97.6	97.9	97.8
Inv-Net (3)	95.2	94.7	96.0	95.0	97.1	96.8	97.4	97.1
Inv-Net (4)	97.7	98.5	97.8	97.7	96.6	97.2	97.9	97.5

as INN of our Inv-Net model. Besides, we also compare our model performance with the existing research methods for the classification task implemented in our FLAME dataset.

B. Segmentation Task

The segmentation performance of Inv-Net is evaluated to established segmentation models such as SegFormer [62], FPN [63], PSPNet [64], U-Net [65], DeepLab V3 [66]. Furthermore, we also compare our model with the existing studies for the segmentation task in our FLAME dataset. However, this implementation involves the integration of numerous state-of-the-art models along with our Inv-Net model.

C. Multitask Learning

We experiment and asses our proposed multitask Inv-Model evaluation against other scholar existing studies and their techniques, including Jaouad et al. [67], DSI-Net [68], URepNet-v3+SVM [69]. Each multitask framework has been trained under our FLAME dataset and performance assessments are conducted entirely on the authorized testing database.

VIII. RESULT AND DISCUSSION

We assess the efficacy of our proposed Inv-Net framework by comparing it to other cutting-edge models and previous research, utilizing defined assessment measures across different configurations in ablation tests. We evaluate its performance in classification, segmentation tasks, and overall multitasking abilities.

A. Ablation Studies

Table VI illustrates an ablation study covering various models for both segmentation and classification tasks, distinguished by diverse setups and methodological methods. Variations in dropout rates, optimization methodologies, batch sizes, and layer-freezing tactics are thoroughly studied. Results reveal that the proposed model consistently outperforms other approaches across segmentation and classification tasks, achieving maximal accuracy, precision, recall, and F1-score metrics. Specifically, the model proposed attains an accuracy of 98.2% for segmentation and 97.9% for classification, highlighting its efficacy. Besides, Table VII outlines performance metrics throughout several iterations of the Inv-Net model, spanning segmentation and classification tasks and encompassing varied configurations of the INN network. Noteworthy is Inv-Net (4), which notably obtains the highest segmentation accuracy of 97.7% and precision of 98.5%, indicating its ability to precisely outline segmented regions. Conversely, in the area of classification, Inv-Net (2) emerges as the top performer, having an accuracy of 97.8% and an F1-score of 97.8%, so demonstrating its ability to correctly categorize data instances. In addition, Fig. 14 exhibits accuracy curves spanning training and testing datasets, providing a detailed study of trends across 100 epochs on different configurations of INN models.

B. Classification Task

The resulting Table VIII presents the performance metrics of various models in a categorization task. Models including DenseNet161, ResNet101, VGG16, DCNN, and Inv-Net were examined. Notably, Inv-Net displays remarkable performance across all parameters, reaching the highest accuracy of 98.1%, precision of 99.2%, F1-score of 98.1%, and AUC of 99.6%. DenseNet161 also displays impressive results with an accuracy of 96.34% and a precision of 96.2%. However, VGG16 demonstrates comparably poorer performance with an accuracy of 94.3% and an F1-score of 94.5%. Besides, in Fig. 15 the confusion matrices are shown where our proposed Inv-Net model generates larger diagonal values compared to other models.

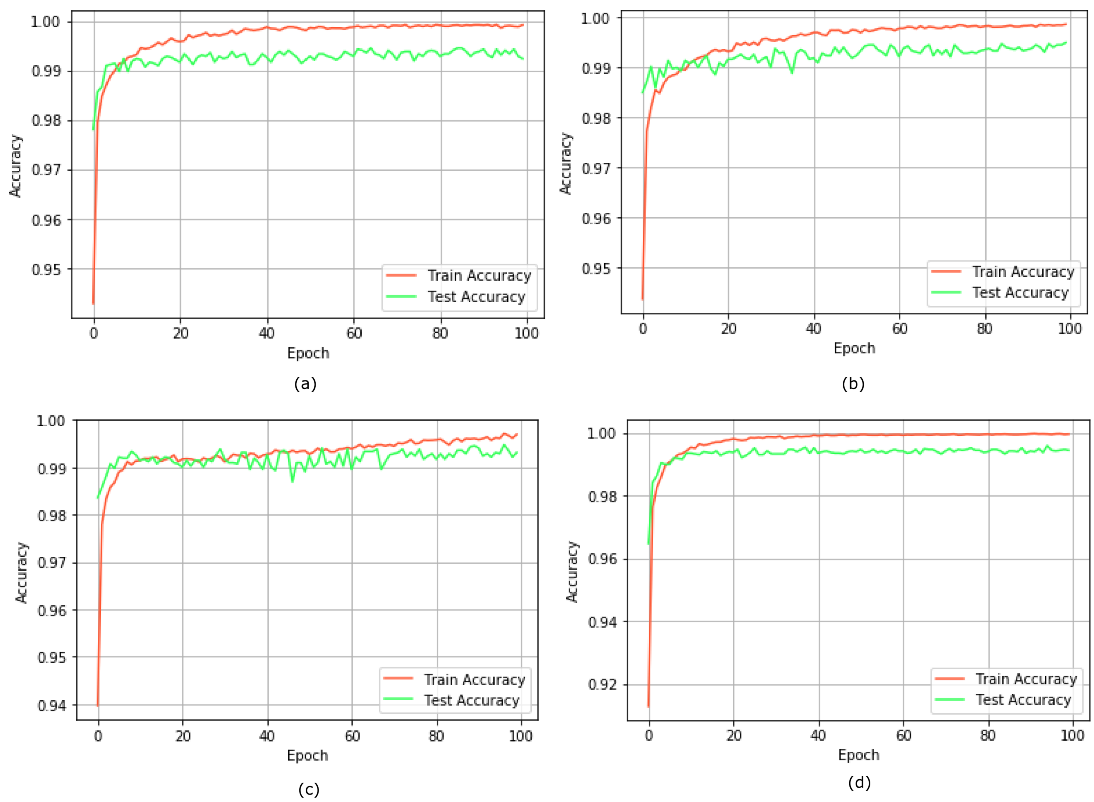

Fig. 14. Ablation Study on training and testing accuracy curves for Inv-Net models: (a) Inv-Net(1), (b) Inv-Net(2), (c) Inv-Net(3), and (d) Inv-Net(4).

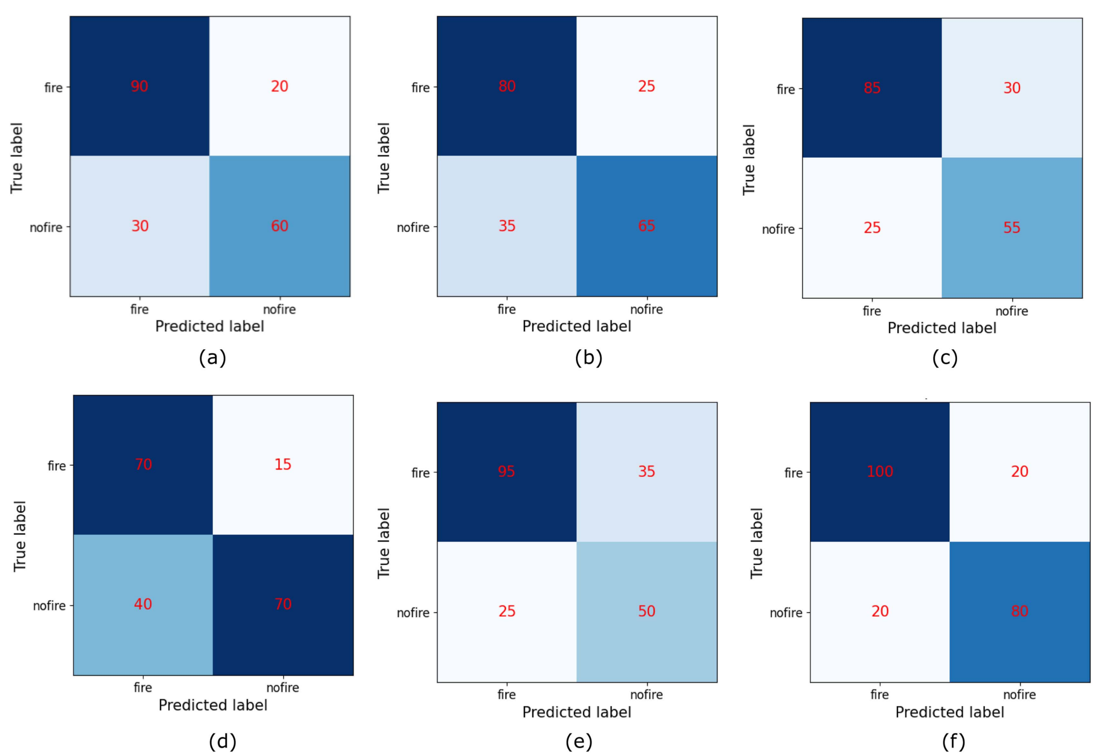

Fig. 15. Confusion matrix results from our Inv-Net model, using (f) INN as the backbone, compared to (a) ResNet101, (b) EfficientNetB4, (c) DenseNet121, (d) VGG16, and (e) DCNN, (f) INN.

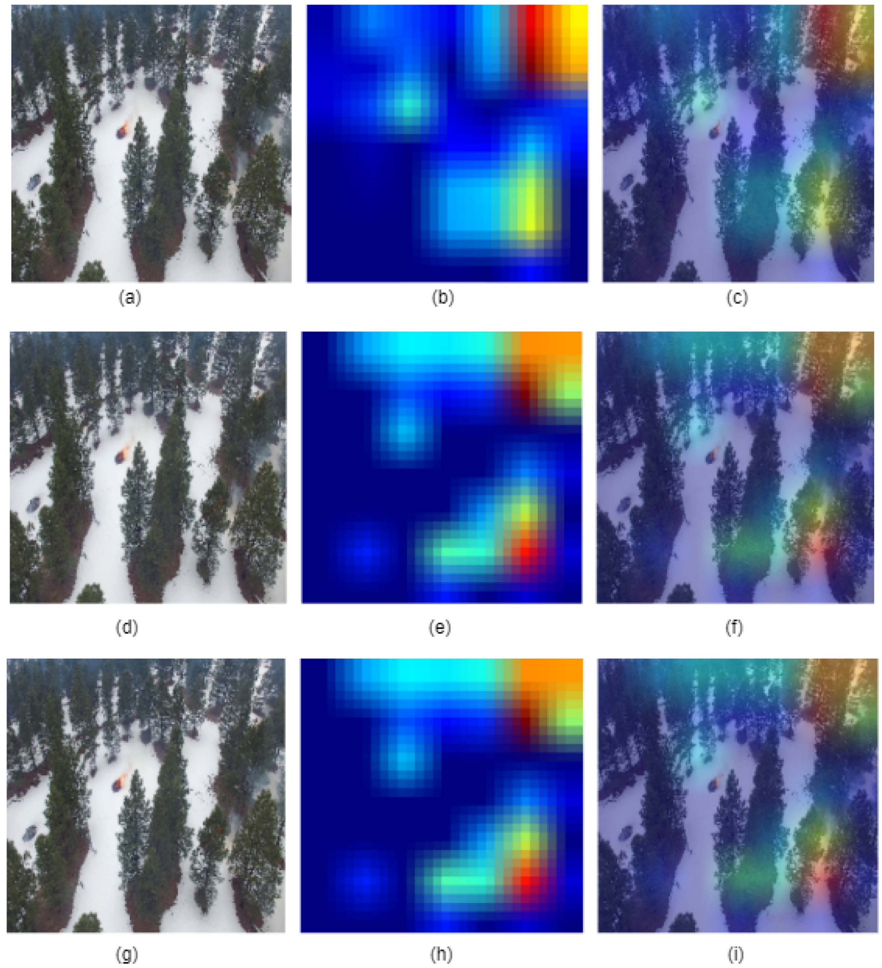

Fig. 16. Ablation study on Inv-Net classification model's performance. Images (a), (d), and (g) show dehazed inputs; (b), (e), and (h) show pixelated versions; (c), (f), and (i) display final attention maps.

TABLE VIII
COMPARISON OF THE MEAN ± STANDARD DEVIATION (S.D.) OF INV-NET PERFORMANCE AGAINST CONVENTIONAL DEEP LEARNING MODELS FOR THE WILDFIRE CLASSIFICATION TASK

Model	ACC	PRE	REC	F1-score
Resnet101	92.31α	$90.2 \pm 6.1\alpha$	$92.4 \pm 1.5\alpha$	90.13α
EfficientNetB4	93.1	93.2 ± 4.1	93 ± 2.3	92.4
DenseNet161	96.34	96.2 ± 5.8	96 ± 3.9	96.9
VGG16	94.3α	$94.1 \pm 4.1\alpha$	$94.2 \pm 5.8\alpha$	94.5α
DCNN	96.12	95.2 ± 8.5	95.4 ± 4.8	95.2
Inv-Net	98.1	99.2 ± 1	97.2 ± 1.1	98.1

Statistical significance is denoted by α at $p < 0.05$ using a two-sided McNemar's test against the proposed Inv-Net strategy.

TABLE IX
QUANTITATIVE COMPARISON OF CLASSIFICATION PERFORMANCE WITH DIFFERENT METHODS TRAINED ON THE FLAME DATASET

Model	ACC	PRE	REC
Dense-MobileNet [72]	95.7	93.2	94.2
CrossViT [73]	93.4	93.2	93.2
DSI-Net [68]	96.8	94.5	93.9
FBC-Anet [74]	90.6	92.4	93.7
Proposed	98.2	99.3	97.5

Table IX, the model we propose has shown significant advantages over other methods in terms of classification performance, according to a recent quantitative comparison conducted on the FLAME Dataset. With an accuracy (ACC) of 98.2%, precision (PRE) of 99.3%, and recall (REC) of 97.5%, our proposed approach surpasses Dense-MobileNet, CrossViT, DSI-Net, and FBC-Anet by substantial margins.

In Fig. 16, we utilize class activation mapping (CAM) [79] to evaluate our classification performance by interpreting regions of fire using the Inv-Net model architecture. For this work, the segmentation component is frozen to assess the classification task's contribution to the study.

C. Segmentation Task

Table X displays the performance characteristics of current state-of-the-art multiple models in a segmentation test, notably, InV-Net beats other models with the highest accuracy of 98.2% and IoU of 97.6%, indicating its superior capacity to reliably segment objects. FPN also displays high performance with an accuracy of 97.3% and an IoU of 98.1%. However, SegFormer and U-Net demonstrate considerably poorer performance with accuracies of 91.3% and 93.2%, respectively. In Table XI,

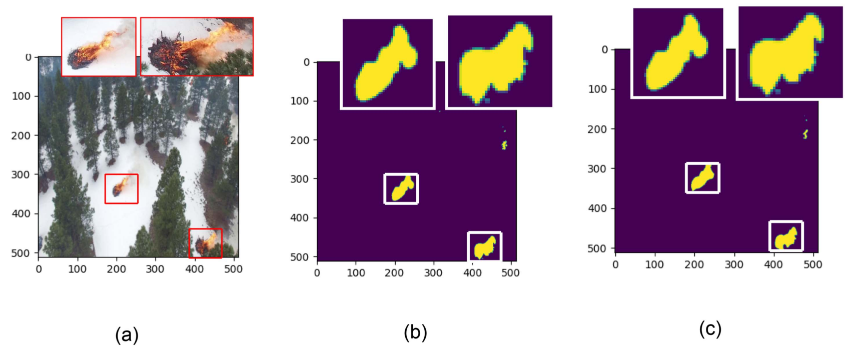

Fig. 17. Segmentation performance of our Inv-Net model on the FLAME Dataset. (a) Input images. (b) Ground truth. (c) Prediction.

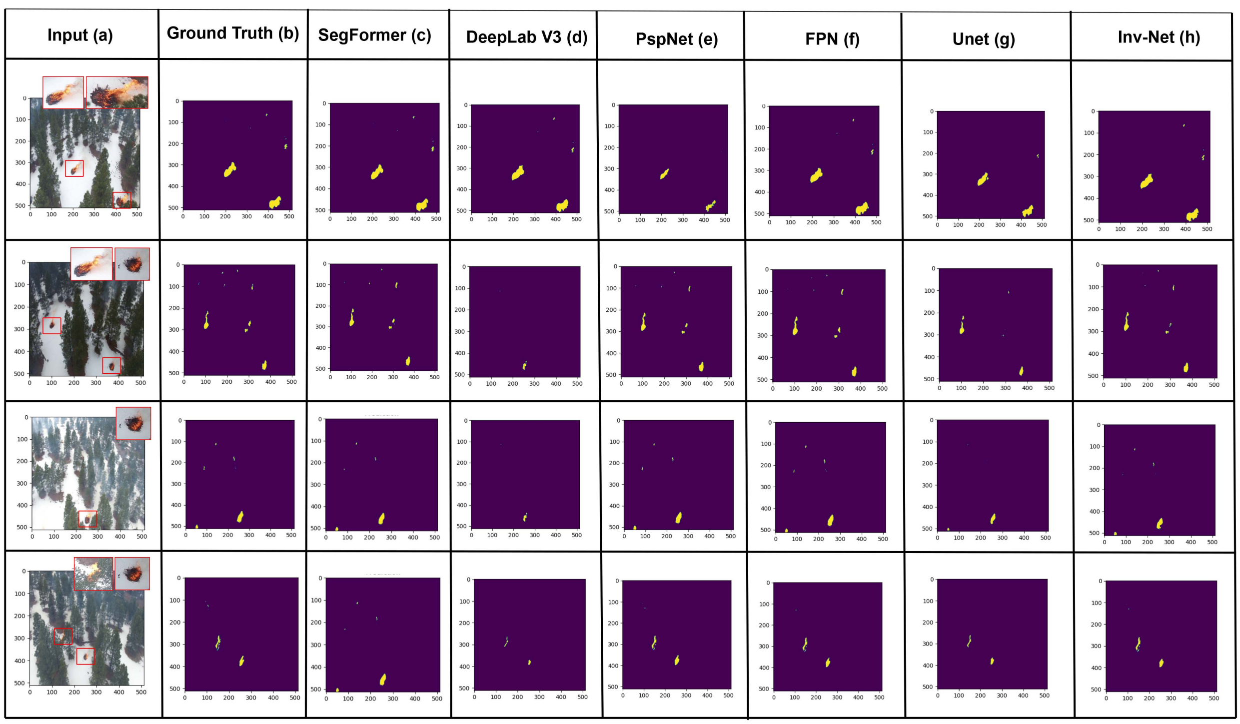

Fig. 18. Segmenation performance on different segmentation models. (c) Segformer. (d) DeepLab V3. (e) PspNet. (f) FPN. (g) U-net with our proposed (h) Inv-Net model comparing with (a) original images, (b) ground truth images.

TABLE X
COMPARISON OF INV-NET PERFORMANCE AGAINST TRANSFER LEARNING MODELS FOR THE WILDFIRE SEGMENTATION TASK

Model	ACC	PRE	REC	IoU
SegFormer	91.3α	$92.1 \pm 5.1\alpha$	$92.2 \pm 2.2\alpha$	91.2α
FPN	97.3	96.1 ± 2	97.1 ± 1.2	98.1
PSPNet	94.9	93.8 ± 5.8	96 ± 3.9	94.9
U-Net	93.2α	$95.2 \pm 7.1\alpha$	$94.2 \pm 5.4\alpha$	95.3α
DeepLab V3	92.1	94.2 ± 7.2	93.2 ± 5.3	93.3
Inv-Net	98.2	97.12 ± 5	98.3 ± 2.3	97.6

Results for the segmentation of each studied fire burn lesion are presented individually, with the overall column showing the average of the individual lesion results (mean ± S.D.). Statistical significance denoted by β at $p < 0.05$, paired t-test against the proposed Inv-Net strategy.

our proposed approach has been evaluated quantitatively with existing studies in terms of segmentation performance using the FLAME Dataset. The results show that our model outperforms other methods, obtaining significantly higher accuracy, precision, and IoU metrics. Our proposed model achieves an accuracy (ACC) of 97.8%, precision (PRE) of 95.3%, and intersection over union (IoU) of 95.7%. It outperforms UT-Net, TransUnet, FBC-ANet, and Axial-DeepLab by considerably. Furthermore, Figs. 17 and 18 illustrate the performance evaluation of our model in recognizing burn lesions within specified locations, alongside comparative assessments with other models.

TABLE XI
QUANTITATIVE COMPARISON OF SEGMENTATION PERFORMANCE WITH DIFFERENT METHODS TRAINED ON THE FLAME DATASET

Model	ACC	PRE	IoU
UT-Net [75]	82.5	80.3	84.6
TransUnet [76]	90.2	91.8	93.2
FBC-ANet [77]	91.9	91.5	90.8
Axial-DeepLab [78]	79.8	82.7	83.9
Proposed	97.8	95.3	95.7

TABLE XII
PERFORMANCE OF THE PROPOSED INV-NET MODEL COMPARED WITH THE EXISTING MODELS USING THE FLAME DATASET

Model	Segmentation				Classification				Inference Time (ms)	FLOP Count
	ACC	PRE	REC	IoU	ACC	PRE	REC	F1-score		
Jaouad et al. [67]	90.5	91.0	89.8	90.2	88.3	88.7	90.5	88.0	902.249	300
DSI-Net [68]	91.3	91.7	90.8	91.0	92.5	92.8	92.2	92.4	910.249	280
URepNet-v3+SVM [69]	88.9	89.3	88.1	88.6	90.2	90.5	89.9	90.1	915.249	350
Shintaro et al. [70]	89.2	89.5	88.8	89.1	90.5	90.7	90.1	90.3	920.249	340
Abderraouf et al. [71]	89.0	89.4	88.3	88.9	90.3	90.6	90.0	90.2	925.249	330
This Proposed Study	98.2	97.9	97.5	97.1	97.9	97.7	98.3	97.9	640.249	250

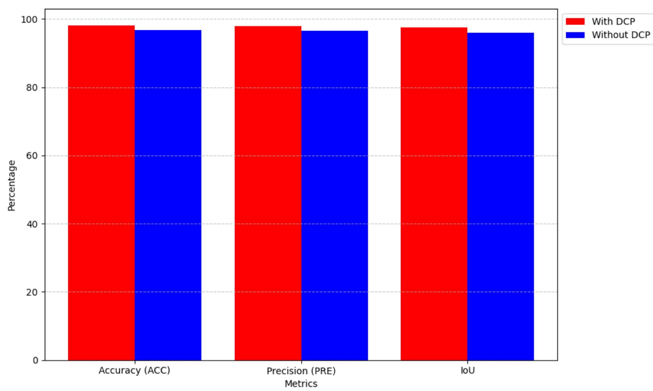

Fig. 19. Performance comparison with and without dark channel prior (DCP).

Notably, our Inv-Net model outperforms alternative techniques, displaying better performance in detecting burn lesions within the test dataset.

D. Multitask Learning

In Table XII, through a comparison analysis of the Inv-Net model and other models in the FLAME dataset, our proposed approach consistently exhibits superior performance in both segmentation and classification tasks. The Inv-Net model demonstrates superior performance compared to the existing models across all tested criteria. Inv-Net outperforms the existing model, DSI-Net, in segmentation with higher accuracy (98.2% versus 91.3%), precision (97.9% versus 91.7%), recall (97.5% versus 90.8%), and IoU (97.1% versus 91.0%). Inv-Net achieves an accuracy of 97.9%, precision of 97.7%, recall of 98.3%, and an F1-score of 97.9% in classification, beating DSI-Net's performance of 92.5%, 92.8%, 92.2%, and 92.4% in the same metrics, respectively. Furthermore, Inv-Net exhibits a reduced inference time of 640.249 ms and a lower computational cost of 250 FLOP Count, indicating a more efficient and effective model in comparison to the other models evaluated. This makes it a superior option for both plant disease detection tasks involving segmentation and classification.

Our study improves wildfire prediction by dynamically adjusting to new data, as opposed to existing static models. This adaptability is achieved by combining multitask learning, INN layers, and improved picture preprocessing with DCP. Experiments show that our methodology surpasses previous methods in terms of accuracy and responsiveness. In addition, the use of the DCP approach improves the performance of our Inv-Net model in comparison to not using it, as illustrated in Fig. 19. Integrating the DCP approach into the multitask framework has led to improved results in terms of accuracy, precision, and IoU score.

IX. CONCLUSION AND FUTURE WORK

Our study proposed an innovative methodology that bridged the gap between the DCP method and the multitask learning framework, Inv-Net, to address issues in wildfire detection and classification, particularly amid atmospheric disturbances. During the initial stage of the image dehazing process, the DCP approach achieved a 22.1% increase in SSIM, a 13% improvement in PSNR, and a 96% reduction in execution time compared to the multiscale retinex method. Subsequently, our proposed Inv-Net model significantly improved segmentation accuracy by 7.6%, classification accuracy by 5.8%, reduced inference time by 30%, and decreased FLOP count by 10.7% in comparison to DSI-Net. Through the seamless integration of these methodologies, we gained enhanced accuracy and efficiency in wildfire burn analysis, substantially developing wildfire control measures. By concurrently training segmentation and classification tasks, our research not only boosted the precision of wildfire detection but also sped the process, unleashing more effective wildfire management strategies. Moving ahead, there were appealing prospects for additional exploration and development. However, despite the considerable gains obtained by our technique, certain restrictions remained. The DCP method may have had issues in reacting to quickly changing climatic circumstances, and the processing needs of deep learning techniques could have hampered real-time deployment in operational wildfire management systems. Future research efforts should have focused on optimizing the efficiency and scalability of our technique to permit practical application in real-world circumstances. Future endeavors could have focused on fine-tuning the Inv-Net design, researching novel data augmentation methodologies, and extending the application of our technology to real-time wildfire monitoring systems. Furthermore, as wildfires continued to pose significant challenges globally, investigating the scalability of our approach to larger geographic areas and diverse environmental conditions was crucial for its broader practical deployment. By expanding the scope of our research, we could have unlocked new insights and solutions that were pivotal for mitigating the impact of wildfires and safeguarding vulnerable ecosystems and communities.

REFERENCES

[1] W. Li, J. Zhu, S. Pirasteh, Q. Zhu, L. Fu, and J. Wu, "Investigations of disaster information representation from geospatial perspective: Progress, challenges, and recommendations," *Trans. GIS*, vol. 26, no. 3, pp. 1376–1398, 2022, Trans. GIS, doi: 10.1111/tgis.12922.

[2] D. Mafi-Gholami, S. Pirasteh, J.C. Ellison, and A. Jaafarid, "Fuzzy-based vulnerability assessment of coupled social-ecological systems to multiple environmental hazards," *J. Environ. Manage.*, vol. 299, no. 2021, 2021, Art. no. 113573, doi: 10.1016/j.jenvman.2021.113573.

[3] B. Zhang et al., "Improved forest signal detection for space-borne photon-counting LiDAR using automatic machine learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1–13, 2023.

[4] J. Lecina-Diaz, J. Martńnez-Vilalta, A. Alvarez, J. Vayreda, and J. Retana, "Assessing the risk of losing forest ecosystem services due to wildfires," *Ecosystems*, vol. 24, pp. 1–15, 2021.

[5] J. S. Bergtold, M. M. Caldas, A. Joslin, and M. Gharib, "Wildfire across agricultural landscapes: Farmer and rancher experiences and perceptions in the southern great plains," *Environ. Hazards*, vol. 23, pp. 1–26, 2024.

[6] M. Sdraka et al., "FLOGA: A machine learning ready dataset, a benchmark and a novel deep learning model for burnt area mapping with Sentinel-2," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 7801–7824, 2024.

[7] Z. Dong, F. Zhao, G. Wang, Y. Tian, and H. Li, "A. deep learning framework: Predicting fire radiative power from the combination of polar-orbiting and geostationary satellite data during wildfire spread," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 10827–10841, 2024.

[8] S. Li and T. Banerjee, "Spatial and temporal pattern of wildfires in California from 2000 to 2019," *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 8779.

[9] S. N. Saydirasulovich, M. Mukhiddinov, O. Djuraev, A. Abdusalomov, and Y. I. Cho, "An improved wildfire smoke detection based on YOLOv8 and UAV images," *Sensors*, vol. 23, no. 20, 2023, Art. no. 8374.

[10] J. G. Pausas and J. E. Keeley, "Wildfires and global change," *Front. Ecol. Environ.*, vol. 19, no. 7, pp. 387–395, 2021.

[11] D. Virgilio et al., "Climate change increases the potential for extreme wildfires," *Geophysical Res. Lett.*, vol. 46, no. 14, pp. 8517–8526, 2019.

[12] T. Feng et al., "Characterizing fire-induced forest structure and above-ground biomass changes in boreal forests using multi-temporal LiDAR and landsat," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 10108–10125, 2024.

[13] H. Li et al., "High-resolution terrain modeling using airborne LiDAR data with transfer learning," *Remote Sens.*, vol. 13, 2021, Art. no. 3448, doi: 10.3390/rs13173448.

[14] S. Xie, M. Zhou, C. Wang, and S. Huang, "CSPPartial-YOLO: A. lightweight YOLO-Based method for typical objects detection in remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 388–399, 2024.

[15] G. Cheng, X. Xie, J. Han, L. Guo, and G. S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, 2020.

[16] W. Liu, K. Quijano, and M. M. Crawford, "YOLOv5-Tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8085–8094, 2022.

[17] K. Chen et al., "Assimilation of SBAS-InSAR vertical displacement into land surface model to improve the estimation of terrestrial water storage," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2826–2835, 2022, doi: 10.1109/JSTARS.2022.3162228.

[18] P. Jain, S. C. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, "A. review of machine learning applications in wildfire science and management," *Environ. Rev.*, vol. 28, no. 4, pp. 478–505, 2020.

[19] C. Raymond et al., "Understanding and managing connected extreme events," *Nature Climate Change*, vol. 10, no. 7, pp. 611–621, 2020.

[20] J. Wang et al., "Detecting nighttime fire combustion phase by hybrid application of visible and infrared radiation from Suomi NPP VIIRS," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111466.

[21] S. P. H. Boroujeni et al., "A comprehensive survey of research towards AI-enabled unmanned aerial systems in pre-, active-, and post-wildfire management," *Inf. Fusion*, vol. 108, 2024, Art. no. 102369.

[22] M. Pecha, Z. Langford, D. Horák, and R. Tran Mills, "Wildfires identification: Semantic segmentation using support vector machine classifier," *Programs Algorithms Numer. Math.*, pp. 173–186, 2023.

[23] A. Akbari Asanjan, M. Memarzadeh, P. A. Lott, E. Rieffel, and S. Grabbe, "Probabilistic wildfire segmentation using supervised deep generative model from satellite imagery," *Remote Sens.*, vol. 15, no. 11, 2023, Art. no. 2718.

[24] R. Ghali and M. A. Akhloufi, "Deep learning approaches for wildland fires using satellite remote sensing data: Detection, mapping, and prediction," *Fire*, vol. 6, no. 5, 2023, Art. no. 192.

[25] Y. Wang, Y. Wang, C. Xu, X. Wang, and Y. Zhang, "Computer vision-driven forest wildfire and smoke recognition via IoT drone cameras," *Wireless Netw.*, vol. 30, pp. 1–14, 2024.

[26] G. Li, P. Cheng, Y. Li, and Y. Huang, "Lightweight wildfire smoke monitoring algorithm based on unmanned aerial vehicle vision," *Signal, Image Video Process.*, vol. 18, pp. 1–13, 2024.

[27] P. Chwalek et al., "Downwind fire and smoke detection during a controlled burn–analyzing the feasibility and robustness of several downwind wildfire sensing modalities through Real World Applications," *Fire*, vol. 6, no. 9, 2023, Art. no. 356.

[28] V. Kalaivani and P. Chanthiya, "A. novel custom optimized convolutional neural network for a satellite image by using forest fire detection," *Earth Sci. Inform.*, vol. 15, no. 2, pp. 1285–1295, 2022.

[29] M. H. B. Azami, N. C. Orger, V. H. Schulz, T. Oshiro, and M. Cho, "Earth observation mission of a 6 U CubeSat with a 5-meter resolution for wildfire image classification using convolution neural network approach," *Remote Sens.*, vol. 14, no. 8, 2022, Art. no. 1874.

[30] D. Rashkovetsky, F. Mauracher, M. Langer, and M. Schmitt, "Wildfire detection from multisensor satellite imagery using deep semantic segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7001–7016, 2021.

[31] Y. Kang, E. Jang, J. Im, and C. Kwon, "A. deep learning model using geostationary satellite data for forest fire detection with reduced detection latency," *GIScience Remote Sens.*, vol. 59, no. 1, pp. 2019–2035, 2022.

[32] A. Shamsoshoara, F. Afghah, A. Razi, L. Zheng, P. Z. Fulé, and E. Blasch, "Aerial imagery pile burn detection using deep learning: The FLAME dataset," *Comput. Netw.*, vol. 193, 2021, Art. no. 108001.

[33] V. Khryashchev and R. Larionov, "Wildfire segmentation on satellite images using deep learning," in *Proc. Moscow Workshop Electron. Netw. Technol.*, Mar. 2020, pp. 1–5.

[34] A. B. Abdusalomov, B. M. S. Islam, R. Nasimov, M. Mukhiddinov, and T. K. Whangbo, "An improved forest fire detection method based on the detectron2 model and a deep learning approach," *Sensors*, vol. 23, no. 3, 2023, Art. no. 1512.

[35] H. Pan, D. Badawi, and A. E. Cetin, "Computationally efficient wildfire detection method using a deep convolutional network pruned via Fourier analysis," *Sensors*, vol. 20, no. 10, 2020, Art. no. 2891.

[36] G. Zhang, M. Wang, and K. Liu, "Deep neural networks for global wildfire susceptibility modelling," *Ecological Indicators*, vol. 127, 2021, Art. no. 107735.

[37] H. Pan, D. Badawi, X. Zhang, and A. E. Cetin, "Additive neural network for forest fire detection," *Signal, Image Video Process.*, vol. 14, no. 4, pp. 675–682, 2020.

[38] L. Wang, H. Zhang, Y. Zhang, K. Hu, and K. An, "A. deep learning-based experiment on forest wildfire detection in machine vision course," *IEEE Access*, vol. 11, pp. 32671–32681, 2023.

[39] A. V. Jonnalagadda and H. A. Hashim, "SegNet: A segmented deep learning based convolutional neural network approach for drones wildfire detection," *Remote Sens. Appl.: Soc. Environ.*, vol. 34, 2024, Art. no. 101181.

[40] R. Ghali and M. A. Akhloufi, "CT-Fire: A CNN-Transformer for wildfire classification on ground and aerial images," *Int. J. Remote Sens.*, vol. 44, no. 23, pp. 7390–7415, 2023.

[41] M. Prakash, S. Neelakandan, M. Tamilselvi, S. Velmurugan, S. Baghavathi Priya, and E. Ofori Martinson, "Deep learning–based wildfire image detection and classification systems for controlling biomass," *Int. J. Intell. Syst.*, vol. 2023, no. 1, 2023, Art. no. 7939516.

[42] G. L. James et al., "An efficient wildfire detection system for AI-embedded applications using satellite imagery," *Fire*, vol. 6, no. 4, 2023, Art. no. 169.

[43] S. A. Sifat, M. Hasan, S. B. N. Joy, M. A. Habib, and R. Rahman, "PyroVision: A. deep learning based model for wildfire detection in satellite imagery," in *Proc. 6th Int. Conf. Elect. Eng. Inf. Commun. Technol.*, May 2024, pp. 705–710.

[44] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, no. 1, vol. 2018, 2018, Art. no. 7068349.

[45] L. Alzubaidi et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, pp. 1–74, 2021.

[46] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 270–279.

[47] L. Liu, X. Song, M. Wang, Y. Liu, and L. Zhang, "Self-supervised monocular depth estimation for all day images using domain separation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 12737–12746.

[48] M. A. Finney, S. S. McAllister, J. M. Forthofer, and T. P. Grumstrup, *Wildland Fire Behaviour: Dynamics, Principles and Processes*. Clayton South, VIC, Australia: CSIRO Publishing, 2021.

[49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[50] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[51] A. A. J. Pazhani and S. Periyanayagi, "A novel haze removal computing architecture for remote sensing images using multi-scale retinex technique," *Earth Sci. Inform.*, vol. 15, no. 2, pp. 1147–1154, 2022.

[52] I. Kansal and S. S. Kasana, "Improved color attenuation prior based image de-fogging technique," *Multimedia Tools Appl.*, vol. 79, no. 17, pp. 12069–12091, 2020.

[53] D. Li et al., "Involution: Inverting the inherence of convolution for visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 12321–12330.

[54] Y. Zhou et al., "Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images," *Med. Image Anal.*, vol. 70, 2021, Art. no. 101918.

[55] L. Song, J. Lin, Z. J. Wang, and H. Wang, "An end-to-end multi-task deep learning framework for skin lesion analysis," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 10, pp. 2912–2921, Oct. 2020.

[56] I. R. I. Haque and J. Neubert, "Deep learning approaches to biomedical image segmentation," *Inf. Med. Unlocked*, vol. 18, 2020, Art. no. 100297.

[57] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[58] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[59] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[60] S. Mascarenhas and M. Agarwal, "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for image classification, in *Proc. 2021 Int. Conf. Disruptive Technol. Multi-Disciplinary Res. Appl.*, Nov. 2021, vol. 1, pp. 96–99.

[61] F. Cheng, S. L. Wang, X. Z. Wang, A. W. C. Liew, and G. S. Liu, "A global and local context integration DCNN for adult image classification," *Pattern Recognit.*, vol. 96, 2019, Art. no. 106983.

[62] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 12077–12090, 2021.

[63] S. Seferbekov, V. Iglovikov, A. Buslaev, and A. Shvets, "Feature pyramid network for multi-class land segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 272–275.

[64] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2881–2890.

[65] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.- Assist. Interv., 18th Int. Conf.*, Munich, Germany: Springer International Publishing, Oct. 2015, pp. 234-241.

[66] S. C. Yurtkulu, Y. H. Sahin, and G. Unal, "Semantic segmentation with extended DeepLabv3 architecture," in *Proc. 2019 27th Signal Process. Commun. Appl. Conf.*, Apr. 2019, pp. 1–4.

[67] J. Tagnamas, H. Ramadan, A. Yahyaouy, and H. Tairi, "Multi-task approach based on combined CNN-transformer for efficient segmentation and classification of breast tumors in ultrasound images," *Vis. Comput. Ind., Biomed., Art*, vol. 7, no. 1, 2024, Art. no. 2.

[68] M. Zhu, Z. Chen, and Y. Yuan, "DSI-Net: Deep synergistic interaction network for joint classification and segmentation with endoscope images," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3315–3325, Dec. 2021.

[69] M. Song and Y. Kim, "Optimizing proportional balance between supervised and unsupervised features for ultrasound breast lesion classification," *Biomed. Signal Process. Control*, vol. 87, 2024, Art. no. 105443.

[70] A. Amrani, D. Diepeveen, D. Murray, M. G. Jones, and F. Sohel, "Multi-task learning model for agricultural pest detection from crop-plant imagery: A. bayesian approach," *Comput. Electron. Agriculture*, vol. 218, 2024, Art. no. 108719.

[71] S. Sukegawa et al., "Evaluation of multi-task learning in deep learning-based positioning classification of mandibular third molars," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 684.

[72] W. Wang, Y. Li, T. Zou, X. Wang, J. You, and Y. Luo, "A novel image classification approach via dense-MobileNet models," *Mobile Inf. Syst.*, vol. 2020, 2020, Art. no. 7602384.

[73] C. F. R. Chen, Q. Fan, and R. Panda, "Crossvit: Cross-attention multi-scale vision transformer for image classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 357–366.

[74] K. Dong, C. Zhou, Y. Ruan, and Y. Li, "MobileNetV2 model for image classification," in *Proc. 2nd Int. Conf. Inf. Technol. Comput. Appl.*, 2020, pp. 476–480.

[75] Y. Gao, M. Zhou, and D. N. Metaxas, "UTNet: A hybrid transformer architecture for medical image segmentation," in *Proc. Med. Image Comput. Comput. Assist. Interv., 24th Int. Conf.*, Strasbourg, France, Springer International Publishing, Sep. 2021 pp. 61–71.

[76] J. Chen et al., "TransUNet: Rethinking the U-net architecture design for medical image segmentation through the lens of transformers," *Med. Image Anal.*, vol. 97, 2024, Art. no. 103280.

[77] L. Zhang, M. Wang, Y. Ding, T. Wan, B. Qi, and Y. Pang, "FBC-ANet: A semantic segmentation model for UAV forest fire images combining boundary enhancement and context awareness," *Drones*, vol. 7, no. 7, 2023, Art. no. 456.

[78] Y. Wang, E. J. Choi, Y. Choi, H. Zhang, G. Y. Jin, and S. B. Ko, "Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning," *Ultrasound Med. Biol.*, vol. 46, no. 5, pp. 1119–1132, 2020.

[79] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.