Real-Time Wildfire Localization on the NASA Autonomous Modular Sensor using Deep Learning

Yajvan Ravan*, Aref Malek†, Chester V. Dolph‡, and Nikhil Behari§
NASA Langley Research Center, Hampton, VA, 23681

High-altitude, multi-spectral, aerial imagery is scarce and expensive to acquire, yet it is necessary for algorithmic advances and application of machine learning models to high-impact problems such as wildfire detection. We introduce a human-annotated dataset from the NASA Autonomous Modular Sensor (AMS) using 12-channel, medium to high altitude (3 - 50 km) aerial wildfire images similar to those used in current US wildfire missions. Our dataset combines spectral data from 12 different channels, including infrared (IR), short-wave IR (SWIR), and thermal. We take imagery from 20 wildfire missions and randomly sample small patches to generate over 4000 training and 400 test images with high variability, including occlusions by smoke/clouds, easily-confused false positives, and nighttime imagery.

We demonstrate results from a deep-learning model to automate the human-intensive process of fire perimeter determination. We train two deep neural networks, one for image classification and the other for pixel-level segmentation. The networks are combined into a unique real-time segmentation model to efficiently localize active wildfire on an incoming image feed. Our model achieves 96% classification accuracy, 74% Intersection-over-Union(IoU), and 84% recall surpassing past methods, including models trained on satellite data and classical color-rule algorithms. By leveraging a multi-spectral dataset, our model is able to detect active wildfire at nighttime and behind clouds, while distinguishing between false positives. We find that data from the SWIR, IR, and thermal bands is the most important to distinguish fire perimeters. To advance future work in this domain, our dataset and code will be made publicly available.

I. Introduction

Wildfires are a pervasive and ever-growing problem across the world. With the advent of climate change, wildfires are becoming stronger, more frequent, and more unpredictable. From property damage to poor air quality, wildfires are the cause of many annual problems for those who live in areas prone to them. In 2021, wildfires generated a cost of $15.5 billion dollars in damages and response [1]. Over 1 in 6 homes in the US is at risk of wildfires over the next 30 years [1]. Unfortunately, according to a joint workshop between NASA and the US Forest Service (USFS), numerous challenges are still present in wildfire management, including limited surveillance resources, accurate detection in cloudy conditions, limited data & model fusion, and challenges with integrating new technologies amidst diverse stakeholders which requires simplicity & interoperability [1]. Recent NASA work pursues other challenges presented in [1]. For example, [2] and [3] describe novel methods to estimate smoke emission and characterize smoke composition using emitted fire radiation.

Most current methods for wildfire imaging and localization involve human interpretation to label captured imagery and direct response teams. The National Infrared Operations (NIROPS) component of the National Interagency Fire Center (NIFC) manages wildfire mapping for incident management across the nation [4]. NIROPS relies on manned aircraft with an onboard infrared sensor to gather large-scale, full-fire aerial imagery at altitudes around 10,000 feet and transmit them to ground stations [4, 5]. Trained human firespotters then label these images with the location of the fire perimeter and heat sources, among other things, and this information is often used to direct response teams, oftentimes up to 12 hours after the original flight [4, 5].

The imaging aircraft employed today by the US Forest Service's (USFS) National Infrared Operations (NIROPS) [4, 5] generate medium-altitude (3-50km), wide-terrain, and wide-spectrum remote-sensing imagery. NASA's Ikhana

*OSTEM Internship Program, Langley Research Center
†OSTEM Internship Program, Langley Research Center
‡Aerospace Engineer, Aeronautics Systems Engineering Branch, AIAA Member.
§OSTEM Internship Program, Langley Research Center

Western States Fire Missions (2007–2009) flew the AMS (Autonomous Modular Sensor) at 12-13km and downlinked near-real-time fire products—hotspots, perimeters, and temperature-calibrated imagery—directly to incident command, demonstrating wide-area coverage with 3-50 meter ground sample distance (GSD) and reliable night operation [4, 6]. Today's U.S. fire agencies continue to rely on this modality: the Forest Service's NIROPS program operates high-altitude line-scanner systems (Phoenix/AMS) to map large incidents—most often at night—with geo-rectified TIFFs (file format) and shapefiles delivered before morning briefings; line-scanner scan volumes reach 100k–400k acres/hour, enabling multiple large fires per sortie [7–9]. State programs mirror this playbook: California's Fire Integrated Real-Time Intelligence System (FIRIS, two King Air B200s) provides near-real-time perimeter and hotspot intelligence to California Department of Forestry and Fire Protection (CAL FIRE), California Governor's Office of Emergency Services (CAL OES), and local agencies [10, 11], and Colorado's Division of Fire Prevention and Control (DFPC) Multi-Mission Aircraft (PC-12, 5.5-6km) streams IR/EO mapping to Colorado Wildfire Information Management System (CO-WIMS) for incident use without congesting the tactical airspace [12, 13]. Beyond operations, NASA's newer Compact Fire Infrared Sensing Testbed (c-FIRST) instrument flew January–March 2025 on the B-200 over the Palisades/Eaton fires to identify lingering hotspots, reinforcing the continued value of high-altitude thermal, infrared sensing for response and post-fire assessment [14, 15].

Such a detection and response pipeline can be subject to expert error and could allow the fire to grow in the downtime between flight and response. Automation with computer vision has been explored in the domain of active wildfire detection. Some efforts have involved using algorithms with pre-determined thresholds for the intensities or ratios of various spectral bands [16–18]. However, this approach is strongly prone to false positives [19]. More recent works have explored deep learning [19–22], but these focus on limited domains that are not widely used in current wildfire response missions, sticking to low-altitude imagery, satellite imagery, daytime imagery, RGB spectral data, or offline processing. As a result it is difficult to integrate these solutions into current wildfire operations.

This work is organized as follows. In Section II, we review past work towards solving this problem. In Section III we explain existing challenges in past work and describe our dataset and imagery. We then describe the neural networks that comprise our model, their training, evaluation, and combination. Finally, in Section IV we explain the results of testing our model, benchmark it against previous work, depict examples of simulated real-time inference, and compare it with ablated versions of our model using less spectral data. With this work, we hope to demonstrate a proof-of-concept of an automated wildfire localization system that can be integrated alongside humans into the aforementioned response pipeline and spark future work in this domain to eventually automate this process entirely. Our main contributions are as follows:

1) We create a **manually-annotated** dataset that mirrors modern-day wildfire observation missions using flights taken with the NASA Autonomous Modular Sensor (AMS) [4]
2) We train classification and segmentation networks on our dataset and combine these into a two-tier wildfire localization model. We demonstrate our algorithm's ability to run in real-time on a simulated image feed.
3) We show that the usage of IR imagery and hand-labelled data improves the performance, especially recall, of our model compared to past methods.

II. Related Work

A. Datasets

Given the niche domain of this problem, very few datasets exist. This presents a sizeable obstacle to training deep learning models for this problem. A primary dataset is the Corsican Fire Database [23] which contains 500 RGB images and 100 multispectral images, of which only a few are aerial images of the type that would be taken from unmanned aerial vehicles (UAVs). Another dataset is the Visifire dataset [24] which contains seven videos of wildfires from an aerial point-of-view, but does not contain annotations. Pereira et. al created a custom dataset using Landsat-8 imagery and color-rule algorithms from [16–18] as ground-truth, along with a 1000 image manually annotated dataset consisting only of positive examples. The other works mentioned above used either the Corsican Fire Database or custom, unpublished datasets extracted from their own experiments or the Internet [19]. Beyond Corsican and Visifire, several public resources target aerial or multi-sensor wildfire imagery. AIDER (Emergency Response) focuses on aerial disaster scene classification; it includes fire/smoke along with floods, collapsed buildings, and traffic accidents (320 fire/smoke images; 2,545 total including 1,200 "normal"), with an expanded AIDERv2 later released at scale [25, 26]. FLAME provides UAV-collected video and frames of prescribed pile burns, including RGB and multiple thermal palettes, with labeled frame sets for classification (39,375 train / 8,617 test) and 2,003 RGB/mask pairs for

segmentation [27]. FLAME-2 extends this idea to paired, time-synchronized RGB/LWIR from a DJI Mavic 2 Enterprise Advanced, releasing 53,451 paired frames for classification extracted from seven video pairs with four frame-level classes (flame/smoke combinations) [28]. FASDD aggregates over 100,000 fire/smoke images across three sub-datasets (ground cameras, UAV, and remote-sensing), with bounding boxes and a georeferenced RS subset—useful for cross-domain detection and small-target scenarios [29]. This dataset is the most similar to ours. However, it contains only low-altitude airborne images and spaceborne images, without any data in between. Despite its abundance, satellite imagery differs distributionally from airborne remote-sensing data, and wildfire response missions favor airborne platforms because they can be directed and timed as needed. Numerous other datasets that are similar have emerged in recent years both with satellite imagery [30] and relatively lower altitude UAV imagery [31–37].

B. Image Segmentation

Image classification and image segmentation are related but distinct: classification assigns a single label to an image, whereas semantic segmentation assigns a class label to each pixel [38]. Progress over the past decade has been driven by deep learning and large benchmarks such as PASCAL VOC [39] and MS COCO [40], which span diverse object categories and scenes. Early deep-learning methods adapted CNN classifiers into fully convolutional networks (FCNs) for dense prediction [41]. Subsequent work improved context capture via spatial/pyramid pooling and dilated (atrous) convolutions [42, 43], and via encoder–decoder designs. A canonical example is U-Net, which uses skip connections to fuse fine and coarse features [44], and has been effective across medical, aerial, and satellite imagery [19, 45]. Another widely used encoder–decoder is DeepLabv3+, which augments atrous spatial pyramid pooling with a lightweight decoder to sharpen boundaries, yielding strong results on PASCAL VOC and Cityscapes [46, 47].

More recently, transformer-based models have set the pace by modeling long-range dependencies with self-attention. Representative families include: (i) pure/ViT-based encoders such as SETR and Segmenter, which pair a transformer encoder with simple upsampling or a mask transformer decoder, achieving competitive accuracy on ADE20K and Cityscapes [47–50]; (ii) hierarchical transformers such as SegFormer, which uses a multi-scale MiT encoder and an MLP decoder to balance accuracy and efficiency on ADE20K/Cityscapes [51]; and (iii) mask-classification transformers such as Mask2Former, which unify semantic, instance, and panoptic segmentation with a transformer decoder over mask embeddings, attaining state-of-the-art results on COCO, ADE20K, and Cityscapes [40, 47, 50, 52]. For instance segmentation, transformer approaches build on or surpass classical baselines like Mask R-CNN [53]. Together, these encoder–decoder CNNs and modern transformer architectures define today's spectrum of high-performing segmentation methods.

C. Color-Rule Algorithms

The problem of wildfire detection is a simplified version of semantic image segmentation, where we wish to isolate which pixels correspond to fire in an image. The simplest solution to this problem relies on identifying fire by color, i.e. using color-space segmentation based on the available spectral channels [54]. For example, Murphy *et al.* [17], Schroeder *et al.* [16], and Kumar and Roy [18] developed color-value rule-based algorithms for the images captured by the LANDSAT-8 satellite [55]. These algorithms rely on techniques such as statistics in small neigbhorhoods around each pixel, reflectance ratios, and band saturation. An example rule from Schroeder *et al.* is shown in Figure 1 where ρ_i is the reflectance in channel i, and $R_{ij} = \rho_i/\rho_j$. However, these algorithms are strongly prone to false positives from objects with similar colors to fire and have limited robustness in real-world applications [54].

$$\{(R_{75} > 2.5) \text{ and } (\rho_7 - \rho_5 > 0.3) \text{ and } (\rho_7 > 0.5)\} \text{ OR}$$
$$\{(\rho_6 > 0.8) \text{ and } (\rho_1 < 0.2) \text{ and } (\rho_5 > 0.4 \text{ or } \rho_7 < 0.1)\}$$

Fig. 1 Algorithm for identifying unambiguous fire pixels according to Schroeder *et al.* [16]

D. Deep Learning Approaches to Wildfire Detection

More promising results come from the application of deep-learning methods to this problem. Lee *et al.* demonstrated classification networks based on a variety of aforementioned architectures [56]. The authors trained on around 20,000 images extracted from aerial videos from UAVs and were able to achieve up to 99% accuracy of classification on fire and

non-fire images [56]. Zhao *et al.* created the network architecture of FireNet, based on AlexNet, and demonstrated 98% classification accuracy of fire and non-fire images [22]. The authors also augment their dataset using saliency-based localization methods to isolate regions of interest containing fire [22]. Park *et al.* explored the use of transfer learning with various architectures for multi-label classification of wildfire images [57].

Much of the aforementioned work has focused on classification or isolation of regions of interest with fire. Frizzi *et al.* trained a network based on the VGG16 architecture to segment RGB images with fire and smoke classes and were able to achieve 72% IoU for fire [58]. Lee *et al.* train a segmentation model based on the mobilenetv3 architecture to perform real-time (59 FPS) fire segmentation on RGB images in [59]. Behari *et al.* train a variety of network architectures for fire segmentation creating ground truth fire labels from Landsat-7 imagery by computing the difference normalized burn ratio (dNBR) index and taking pixels with a dNBR value of over 0.27 as active fire [60]. Pereira *et al.* generated ground truth masks of wildfire imagery taken with the Landsat-8 satellite [55] using color-rule algorithms from [16–18]. They then trained a U-Net based segmentation model with this data, experimenting further with using all 10 spectral channels vs. only 3, achieving up to 81% accuracy on manually annotated data [19]. Finally, Perrolas *et al.* the authors propose a search algorithm using the Quad-Tree method for fire localization [21]. Their algorithm recursively divides a larger image into quadrants if they contain fire as determined by their classification network based on SqueezeNet [61]. Then, once the patches are small enough, they perform segmentation using a U-Net based model. The authors propose a system that takes RGB images from an aerial view and transmits them to a processing unit to run their algorithm offline.

In addition to various segmentation methods, many works have explored variations of object detection techniques. In [62] the authors demonstrate the use of the YOLOV3 [63] object detection algorithm to draw bounding boxes around regions of fire on RGB images with 83% accuracy. In [64], Zhao *et al.* compared various object detection methods such as YOLOV3 and Faster R-CNN showing that they were better at detecting small fire targets but that they are prone to falsely detecting fire-like and smoke-like targets. In [65], the authors combine color-space rules, texture rules, and deep learning to extract features and perform fire segmentation on image patches.

Recently transformers have become more common in the field of machine learning and computer vision, especially with the advent of vision transformers. For example [66], the authors use a network architecture referred to as TransUNet to perform fire segmentation. They train their models using the Corsican Fire Database [23] achieving better performance compared to other deep learning methods.

III. Methodology

In this work, we try to improve upon the existing literature in a few different ways. Firstly, we focus on the domain of high-altitude (3-50 km) remote sensing imagery, specifically using imagery obtained from the Ikhana flights [4], and create a large manually annotated dataset. Airborne thermal and IR imaging has clear operational precedent and advantages for wildland fire intelligence. This data is similar to the medium-altitude, wide-terrain imagery generated by the imaging aircraft employed today by the NIROPS and the USFS [4, 5], but such a human-labelled dataset is not publicly available. These images exhibit a different pattern than the data from aforementioned work such as [23, 58], particularly in that active wildfire is a very small fraction of the image (~2% on average). Compared to satellites, high-altitude airborne sensors occupy a useful "middle ground." Satellites (e.g., MODIS/VIIRS) offer global persistence but are constrained by orbit schedules, cloud/smoke/terrain obscuration, and coarser native resolution (375 m for VIIRS I-band, 1 km MODIS), which limits perimeter fidelity and small fire detection—especially under mixed-pixel conditions [67, 68]. Compared to low-altitude UAVs, high-altitude line-scanners sacrifice centimeter-scale detail but deliver orders-of-magnitude higher scan rates and broad coverage not limited by Temporary Flight Restrictions (TFR); UAV/EO-IR "ball" systems excel for tactical overwatch and spot-fire confirmation but are endurance/FOV-limited and slower for whole-incident mapping [7, 69]. Finally, AMS-style systems also provide dual-gain thermal channels (high-gain sensitivity for subtle contrasts; low-gain headroom for 800–1000 °C), letting a single overflight capture both faint residual heat and active flame fronts without saturation—precisely the regime of interest for our dataset's high-altitude imagery [4, 6].

Secondly, our paper makes use of all available spectral data including infrared (IR), thermal, and short-wave IR in addition to the visual spectrum (RGB). Previous works primarily utilize the visual spectrum (notable exception for [19]) as data input primarily because such images are easily accessible, unlike imagery containing IR or SWIR [21, 58, 66]. However, this leaves room for improvements when vision is obscured by nighttime, clouds, fog, heavy fire smoke, among other obstacles or easily confused scenarios, such as red/orange objects or backgrounds. Despite these benefits, no large, manually labelled dataset containing IR imagery is publicly available to our knowledge. One possible reason is

Dataset	GSD (m/pixel)	Altitude	Spectral Range	Images
AMS (Ours)	3-50m based on alt.	**Aerial (13.5km)**	**Blue - Thermal**	5364
Corsican Fire Database [23]	Ground	Ground	Blue - AMS Band 7	500
Visifire [24]	Ground	Ground	Blue - Red	2684
FASDD [29]	<1m or 10-30m	Ground/UAV/Satellite	**Blue - Thermal**	> 100,000
Pereira Landsat [19]	30m	Satellite 705 km	Blue - SWIR	150,000 [55]
Frizzi *et al.* [58]	Ground	Ground	Blue - Red	8784
Perrolas *et al.* [21]	1-5m	**Aerial** + Ground	Blue - Red	820 + 500
FLAME 2 [28]	< 1m	UAV ~60 m	**Blue - Thermal**	53,451
AIDER [25, 26]	< 1m	UAV	Blue - Red	2,545

Table 1 Comparison of our dataset vs. other wildfire datasets. Public, human-annotated pixel-wise segmentation datasets in this table include AMS (ours), Corsican Fire Database, VisiFire (via the 2022 manual segmentation release), and FLAME-2. FLAME-2 provides paired RGB–LWIR imagery with masks for a 700-image subset, and Active Fire (Landsat-8) also includes a manually annotated subset. FASDD and AIDER do not provide pixel-wise masks. Frizzi *et al.* and Perrolas *et al.* don't provide human annotations. Our dataset is among those with the widest spectral range and is the largest aerial mid-to-high altitude manually-annotated set.

Band	Label	Wavelength, μm	Features
2	Blue	0.45-0.52	Water
3	Green	0.52-0.6	Vegetation
5	Red	0.63-0.69	Heat Spots
9	SWIR II	1.55-1.75	Smoke/Heat
10	SWIR	2.08-2.35	Heat Spots
11	Infrared(IR)	3.60-3.79	Heat Spots
12	Thermal	10.26-11.26	Smoke/Heat

Table 2 Visual Interpretation of the Frequency Bands captured by the AMS

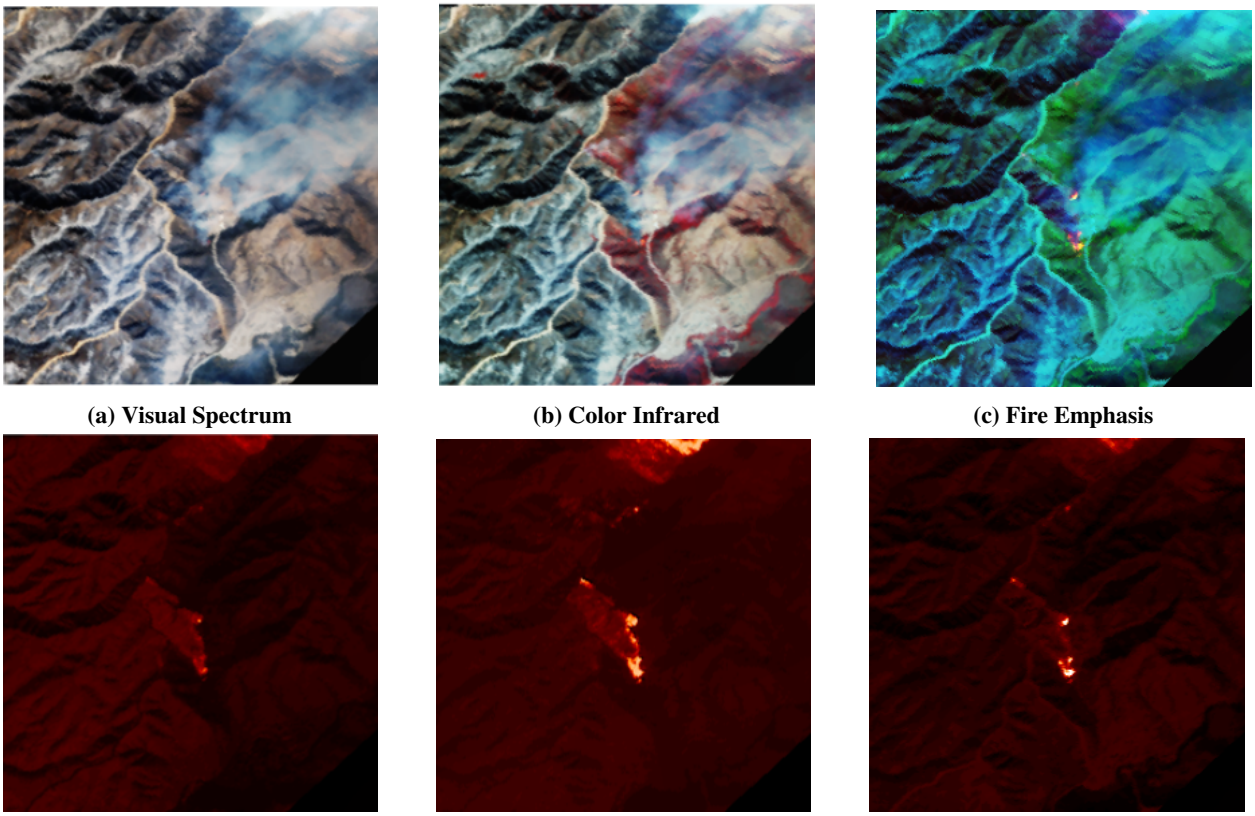
that IR imaging, like in NIROPS missions, must be done with specialized and expensive sensors. Furthermore, some systems may be of military origin, requiring export controls on the data.

Third, we demonstrate a simple real-time segmentation model. Our algorithm uses two neural networks, a classification network and a segmentation network. The former decides whether the current image contains fire and the latter localizes the fire if determined necessary by classifier. The usage of classification prevents useless segmentation, improving speed further. Our models run on patches of size 256×256 pixels. To simulate real-time imaging by an aircraft, we split our test images into a grid of fixed size patches, and input them into our model at the same rate that they were acquired according to flight logs. This patching method allows us to expand our dataset drastically as explained in Table III.A. Furthermore, this ensures that our model is scalable (important for new wildfire technology according to [1]), continuing to operate even when mission parameters such as altitude may generate different outputs.

A. Dataset Preparation

Given the limited availability of human-annotated datasets in this domain, we decided to create our own to avoid the issue of false positives present in the aforementioned color-rule algorithms. A comparison of our dataset with other similar datasets is presented in Table 1. It has the widest spectral range compared to previous datasets, contains the widest variety of difficult situations (occlusion by clouds, nighttime, and easily-mislabelled false positives). Furthermore, of the few datasets that contain aerial imagery, it is the only manually annotated one with over 1000 images.

We used 18 different observation missions completed by the NASA AMS sensor [4] ranging from 2006 to 2019 and in different times of day, smoke levels, and environments. Each image was captured over a target radius of typically 15-nm (equivalent to 1000 mi^2) at an altitude of 13.5km [4]. These images contained reflectance from 16 spectral



(a) Visual Spectrum **(b) Color Infrared** **(c) Fire Emphasis**

(d) Thermal (12) **(e) IR (11)** **(f) SWIR (10)**

Fig. 2 Example of Different AMS Spectral Bands From Our Dataset. (a) Bands Red/Green/Blue, (b) Bands 7/Red/Green, (c) Bands SWIR/7/Red, (d) Band 12 (Thermal), (e) Band 11 (IR), (f) Band 10 (SWIR). Active fire spots are clearly visible in c-f, but obscured in a-b.

bands. We chose to include only bands 1-12, since bands 13-16 are simply low-gain versions of bands 1-12 (with less temperature resolution). Of particular importance to our work were the bands outlined in Table 2. With this spectral data, different visualizations of a particular image are possible. An example of one image is shown in Figure 2. Figure 2a shows the visual spectrum image using the red, green, and blue bands. Figure 2b uses a visualization from [60] called Color Infrared, using band 7, red, and green. This visualization highlights healthy vegetation in red [60]. Figure 2c uses another visualization from [60] called Fire Emphasis, using short-wave infrared (SWIR), band 7, and red. This visualization highlights active fires and hotspots in red [60]. Figures 2d, 2e, and 2f show the infrared (IR), thermal, and SWIR bands independently. In this particular example, the location or presence of fire is not clear with visual information alone, but with the presence of the heat-spotting bands (Thermal, IR, SWIR), it is clear.

Preprocessing. The AMS stores calibrated radiance values that are scaled and accompanied by extensive metadata including solar irradiance, solar zenith angle, temperature correction coefficients, emissivity, and calibration coefficients. The sensor records raw digital numbers (counts) in each band, which are linearly converted to radiance through comparison with blackbodies or ambient references. These radiance values are then scaled due to bit depth constraints and stored alongside scaling factors. For thermal channels (band 11, 12), radiance is calibrated using onboard blackbodies and onboard preprocessing yields brightness temperature values (in K) [70–73].

For non-thermal channels (bands 1-11), we normalize the calibrated radiance by the band-averaged solar irradiance. This approach approximates top-of-atmosphere (TOA) reflectance but omits solar zenith angle correction. This normalization serves multiple purposes: it standardizes channels relative to each other, accounts for slight variations in sensor positioning and lighting conditions, and preserves the visual distinction between daytime and nighttime imagery (nighttime images appear nearly black, as the eye would perceive them). The solar irradiance remains approximately constant across time and space for each band, providing near-constant normalization with minor corrections for lighting

and sensor variation. We clip these normalized values between 0 and 1. While fire adds radiance to lower wavelength channels and can exceed unity, clipping does not significantly impact performance as the fire signal is captured primarily in the heat-sensitive bands (9-12). For thermal channels (12), we convert radiance to brightness temperature and clip values between 250 and 500 K, as temperatures exceeding this range provide diminishing utility for fire detection and normalization. Finally, similar to [60] we resample raw flight data to images with a consistent resolution of 10 m per pixel. This preprocessing pipeline is readily applicable to similar multi-spectral remote sensing platforms. We use 18 flights from the AMS for training and for testing, we generate another dataset from two flights using a similar remote-sensing instrument, the MODIS/ASTER airborne simulator (MASTER) [73], by choosing the 12 spectral bands that overlapped the most in frequency with the AMS.

Patching Augmentation Due to the high cost of aerial missions, aerial imagery datasets tend to be small. However, due to the large area covered, we can generate 250-350 non-overlapping patches of size 256×256 pixels from each image. Different patches from the same image can look very different in terms of terrain as shown in Figure 3b. The small size of images also forces our model to learn the presence of fire with minimal context. The latter is particularly important as wildfire observation aircraft scan laterally and generate pieces of the image in series [4], analogous to patching of the larger image. This resulted in 4875 patches for training and 489 for testing.

Labelling Each image was manually labeled by human inspection. To do so, all available spectral bands were used to mark, by hand, the location of active fire pixels on each image. An example of ground-truth label is shown in Figure 5. Only approximately 18% of the patches contain more than 0.5% active fire pixels.

B. Networks

The architecture of our classification network is shown in Figure 4a. Inspired by the encoder portion of U-Net [44], it contains three layers of down-convolution (with ReLU activation) followed by batch normalization and max pooling. The final two layers are a global max pooling layer and a dense layer with sigmoid activation. The architecture of our segmentation network is shown in Figure 4b. Inspired by the segmentation network in [21], it contains two layers of convolution and downsampling followed by two layers of deconvolution. We use skip connections between the encoder and decoder to maintain details from the high resolution images.

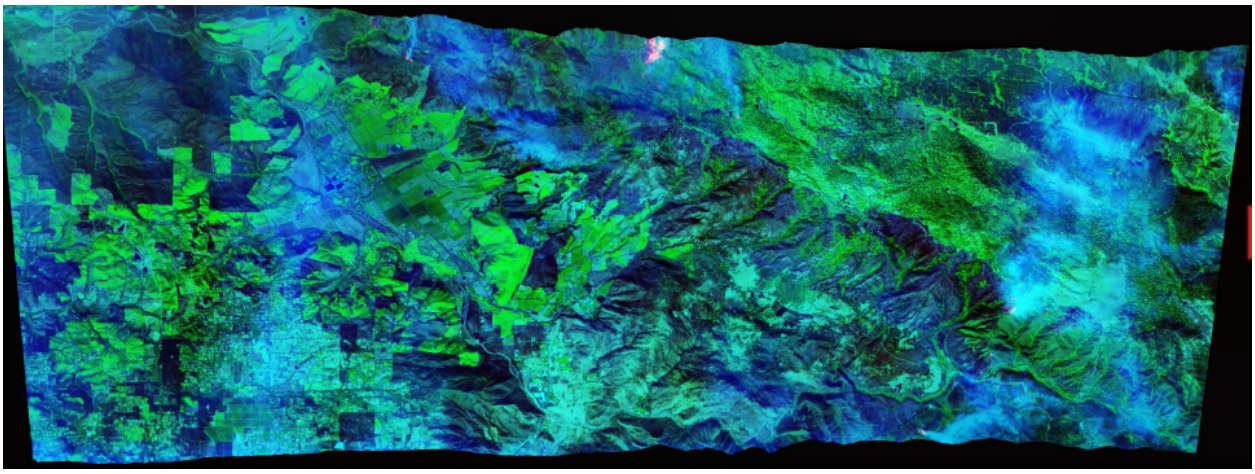
C. Training and Evaluation

The classification network was trained for 500 epochs using the binary cross-entropy loss function and the Adam optimizer [74] using an initial learning rate of $7.5 * 10^{-4}$. We repeatedly sample positive examples to account for class imbalance. The segmentation network was trained using a weighted binary cross-entropy loss and the Adam optimizer [74] with an initial learning rate of $3 * 10^{-4}$. Only positive examples were used to train the segmentation network, with a higher weight of 50 to balance the small fraction of the positive pixels (2%). We evaluated our networks' performance based on accuracy, precision, recall, and IoU for segmentation. Because we intend for these networks to be used in real-time, we also evaluated the inference time for each network.

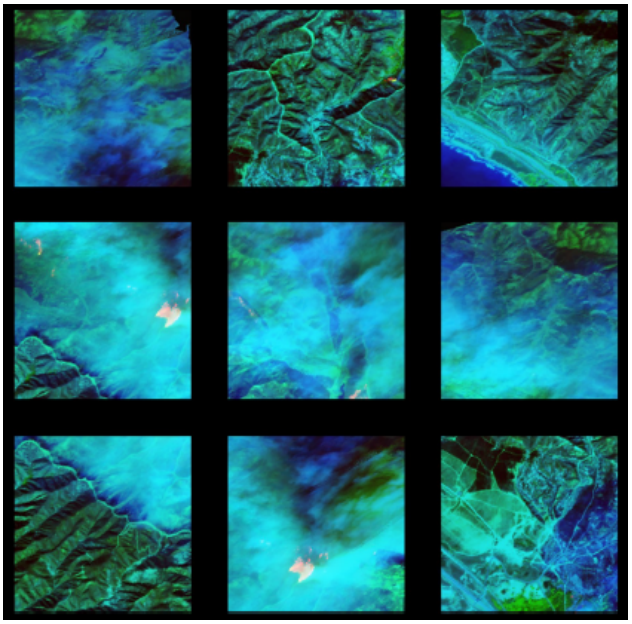
D. Two-Tier Model

We combine these models into a real-time algorithm for wildfire localization. A schematic of our model is shown in Figure 4. For a given patch of size 256×256 pixels, the classifier network first determines if there is fire or not. If there is no fire in the patch (as is the case with 82% of our dataset), then there is no need to run the larger segmentation network, saving computation time. If there is fire, then the segmentation network localizes it. On a static image, we split the image into a grid of patches and run inference on each one-by-one to simulate a real-time feed from an imaging aircraft.

We choose to create such a two-tier model for speed. The segmentation network contains 500 times the number of parameters in the classification network, drastically increasing its inference time (7×, as shown in Table 3). As aforementioned, the large flight paths and high-altitude of wildfire missions results in images that contain little fire, and this two-tier model prevents needless computation. When combined with the image patching approach in our dataset, this allows our model to operate faster on a real-time image feed. Furthermore, the partial interpretability offered by the two-tier system allows for cooperation with human experts as this model is integrated into wildfire missions. For example, the initial deployment may involve using only the classification network, allowing experts to ignore irrelevant portions of an image but still allowing for human-labelling of patches containing wildfire. Such integration can also be used to fine tune either network, while still saving time.

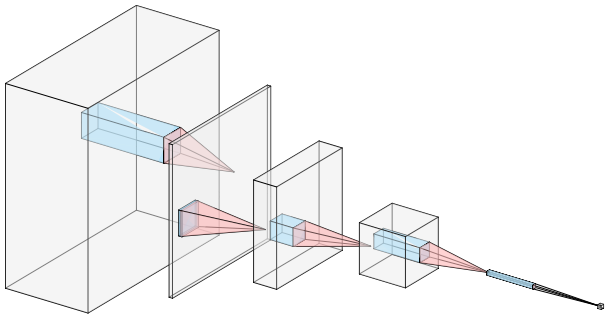


(a) This image is from 1 of 11 California wildfires imaged on October 25, 2007

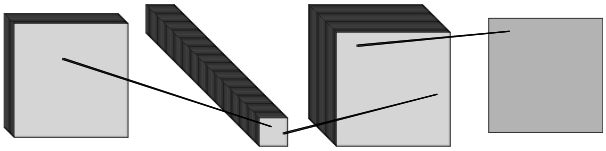


(b) Randomly sampled patches drawn from the above. These patches differ greatly in smoke coverage, vegetation, human development, elevation, etc. This also minimizes the context forcing our model to learn fire detection with minimal information.

Fig. 3 Demonstration of patching on AMS dataset



(a) Classification Architecture: 3-layer encoder from 256×256×12 image to final binary output. Each layer uses Conv + Batch-Norm + Max-Pool.



(b) Segmentation Architecture: 2-layer encoder + 2-layer decoder from 256×256×12 input to 64×64×128 encoding and 256×256×1 output.

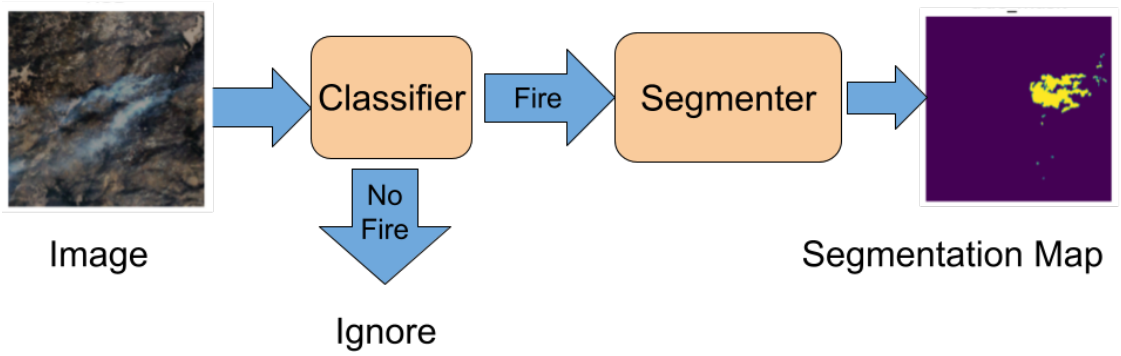


Fig. 4 Inference model combining classification and segmentation networks. Classification detects fire presence; segmentation localizes fire only when needed.

Although more complex compared to the color-rule algorithms, the usage of deep learning allows for greater accuracy, precision, and recall, as explained in Section IV. However, we maintain as much simplicity as possible by keeping our networks shallow and using encoder-decoder segmentation techniques.

IV. Results and Discussion

Network	Acc. %	Prec. %	Recall %	IoU %	Inf. Time
Our Classifier	96.8	82.1	77.8		1.19 ms
Our Segmenter	96.0	58.6	84.0	74.0	7.75 ms
Pereira *et al.*'s segmenter	95.5	53.2	74.8	47.5	

Table 3 Results on the AMS test dataset, measuring accuracy, precision, recall, IoU, and inference time. Our model outperforms Pereira *et al.*'s [19] in all metrics, likely due to the use of human-annotated data and IR spectral data.

Although applying deep learning methods to high-altitude aerial imagery has not been thoroughly explored, Pereira *et al.* opened this exploration by constructing unique deep neural networks for wildfire segmentation on satellite imagery [19]. As the closest approximation, we benchmarked our networks' performances against the segmentation network from [19], which we shall refer to as the Landsat network. The Landsat network was trained using Landsat-8 images that were **automatically** annotated by the aforementioned color-rule algorithms [19]. According to the results in [19], this CNN has strictly greater recall, accuracy, and IoU compared to each of the color-rule algorithms from [16–18]. The Landsat network also only makes use of bands 10, 9, and 2, since according to the results in [19], such a network

performs just as well as one using all available spectral data. We measured the accuracy, precision, recall, and IoU (the ratio of the intersection between ground-truth and predicted fire areas and their union, with 1 indicating perfect prediction and 0 indicating no overlap).

The results of our models on our test set are shown in Table 3. Results are averaged over 10 training seeds. The results show that our segmentation network outperforms the Landsat network, and therefore all of the color-rule algorithms as well. Our model surpasses it in precision (58.6% vs 53.2%), recall (84% vs 74.8%) and significantly in IoU (74% vs 47.5%).

We also chose to train our networks using only spectral bands 10, 9, and 2, which we refer to as the trimmed model, as this is directly comparable to the Landsat network. Results are shown in Section IV.B and Table 4, showing that our methods outperform the Landsat network and color-rule algorithms, as measured by IoU (75% vs 47.5%).

This latter comparison suggests that the use of human-annotated data improves network performance when tested on a human baseline, and also displays a distinct domain difference between satellite imagery and the airborne, high-altitude remote-sensing imagery captured by the AMS. The improvement in accuracy is largely due to a drop in the rate of false negatives, as can be seen in the difference in recall. In this domain, a high recall is arguably more important than high precision, as missed fire has significantly more consequences than a false alarm. Given that both our model and the Landsat network are based on similar UNet architecture, the key reason for performance lies in the datasets. As the color-rule algorithms are known to be inaccurate, the choice of manually-labelled data is strongly preferred, since automated labelling generates poor ground truth labels and therefore inaccurate network training.

The full model is comparable with the trimmed model in most metrics, with a slight improvement in IoU between all of the trimmed versions and the full model, which indicates that the use of additional spectral, particularly the thermal and IR bands improves performance. Thus, these bands contain additonal information that can help distinguish potential errors in the short-wave IR band. Evidently, the improvement from the Landsat network is two-fold, both due to additional spectral information and due to the use of human-annotated data.

We present example results, shown in Figure 5. Figure 5a shows a correctly segmented example. In this image, fire is occluded by clouds in the visual spectrum, and thus would not be caught by most of the aforementioned works. However, our model makes use of the IR band to localize the fire. Figure 5b shows another correctly classified positive example. In this one, the active firespots are difficult to decipher among smoke and previously burned terrain, however, our model correctly localizes the fire, including pieces that are incredibly small and on the border.

A. Real-Time Simulation

The inference time of our networks is shown in Table 3. The results indicate that both networks combined can run at a speed of over 100 fps on our NVIDIA T1200 laptop GPUs *. Each flight with the AMS took roughly 5-10 minutes and generated 200-300 data patches of size 256×256 pixels, so our algorithm comfortably meets real-time processing requirements for the image feed considered.

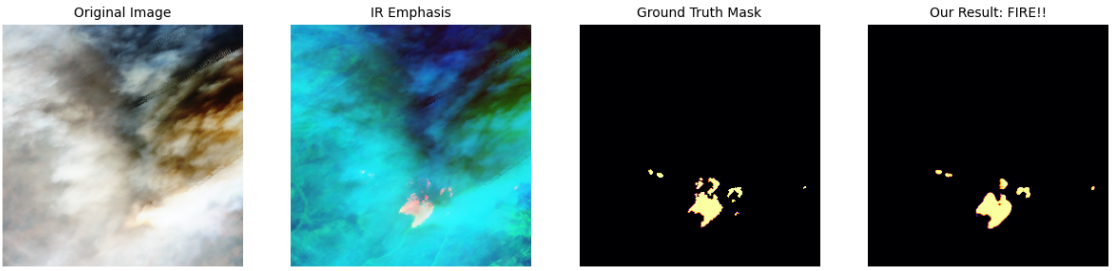
In order to verify this, we validated our algorithm with a simulation of real-time aerial image capture. Images were split into a grid of patches of size 256×256 pixels. These patches were fed in one-by-one from left-to-right and top-to-bottom. Our model then runs inference, as described in Section III.D, and generates 256×256 pixels masks which are either blank, if the classifier determines that there is no fire, or are segmentation output. These masks are laid side-by-side to create our final output. Results showed that the output was generated in negligible time compared to the input feed rate. An example of a partially-completed simulation is shown in Figure 5c for clarification.

B. Spectral Ablation

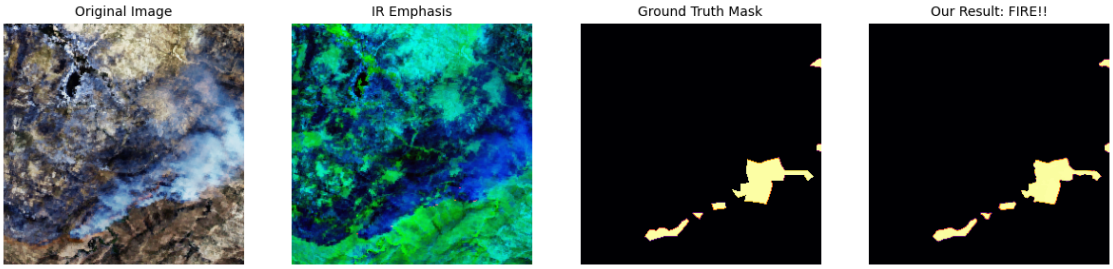
The contribution of various spectral bands to wildfire predictability was also tested by permuting over combinations shown in Table 4. Specific subsets of spectral data were investigated, including the visual spectrum, fire emphasis, SWIR, pure IR, and pure thermal. Separate instances of classification and segmentation were trained using the same dataset and methods above (limited to the respective channel subset). The results are shown in Tables 4 and 5.

First, we note that the model using the RGB image or only the non-thermal bands results in very poor performance. The precision drops to at most 30% for both classification and segmentation and the recall is below 10%. Thus, we can conclude that much of the information needed to determine the presence of fire is contained in bands 9, 10, 11, and 12. Using either bands 9 or 10 achieves similar classification performance compared to the full model as noted by the recall

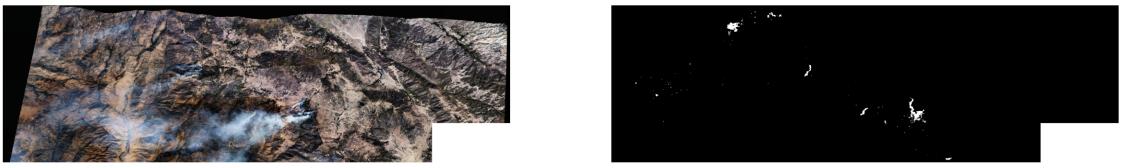
*Trade names or trademarks are used in the report for identification only and that this usage does not constitute an official endorsement, either expressed or implied, by NASA.



(a) In the visual spectrum image, the fire is occluded by clouds, but is readily observable in the IR band.



(b) In the visual spectrum image, the fire is occluded by smoke and vegetation, but the burned area is readily observable in the IR band.



(c) Example of real-time segmentation. The left image depicts the input image being fed in line-by-line (paused halfway through), and the right image depicts the output that our model produced, likewise. White is where fire was detected, and this correlates with simple visual determination based on smoke plumes.

Fig. 5 Example of Results.

Channels	Accuracy %	Precision %	Recall %
11, 9, 2	96.4	78.6	80.0
10, 9, 2	**97.4**	83.3	**84.4**
9	97.1	88.9	74.4
10	95.4	85.3	71.1
11	94.4	66.9	62.2
12	95.8	**97.5**	63.3
1-8	89.8	30.0	4.4
5, 3, 2	89.5	12.2	7.8

Table 4 Classification Results Using Subsets of Spectral Data. High predictability was achieved with bands 10 and 11 only.

and precision, while using either bands 11 or 12 achieves close but notably worse recall (for classification) or precision (for segmentation).

Thus, when limited model size is desirable (e.g. limited memory), one can use a classification network trained on this subset without sacrificing much performance. This also serves to inform that sensors should prioritize imaging in

Channels	Accuracy %	Precision %	Recall %	IoU %
11, 9, 2	95.6	54.9	**79.8**	71.6
10, 9, 2	96.6	64.0	76.8	75.0
9	96.6	**67.8**	60.4	71.7
10	**97.0**	**67.8**	76.6	**76.4**
11	96.4	57.7	66.2	71.5
12	95.8	58.4	63.4	69.5
1-8	94.7	21.5	1.1	47.8
5, 3, 2	94.6	18.4	1.8	48.1

Table 5 Segmentation Results Using Subset of Spectral Data. Performance is good but not comparable to full 12 band model

bands 9, 10, 11, and 12 or similar spectral ranges, as these contain the most important information for wildfire detection. Future work should also explore extracting information from these channels.

V. Conclusion

The contribution of our work is three-fold. We explore the domain of high-altitude remote-sensing wildfire imagery to address the challenges in wildfire localization efforts. Firstly, we create and publish a novel human-annotated dataset derived from wildfires imaged by the NASA AMS sensor [4]. Second, we make use of 12 bands of spectral data to improve accuracy and recall, notably combining SWIR, IR, and thermal channels with the visual spectrum, an approach that, to our knowledge, has seen limited exploration to date in the context of wildfire segmentation. This implies that training data with complete spectral data is more useful than typical RGB images. Finally, we implement a real-time segmentation model for the purpose of scalability, ease-of-use, and efficiency. Future efforts in this area can explore methods for improving the performance of our algorithm. In order for such automation to be deployed in the field, one might desire to minimize false negatives, and hence high recall is desired. Another area to explore is the integration of this algorithm into existing pipelines. This will necessitate data preprocessing, and that step may not be negligible for real-time performance. Furthermore, humans are a necessity in these pipelines until algorithms like ours, become robust and reliable. Therefore, there is a need for exploring human-in-the-loop automation. Finally, as the field of computer vision progresses rapidly, many new techniques arise daily that have yet to be applied to this domain. As wildfires continue to be incredibly costly and deadly worldwide, this problem domain will remain a prime opportunity for automation using computer vision.

VI. Acknowledgements

We wish to thank Anna Trujillo, Robert McSwain, Louis Glaab, Vince Ambrosia, and Marcus Johnson from NASA and Charles Kazimir from NIFC for their help. We would also like to thank the ACERO, STEREO, and TTT projects at NASA for funding this research.

References

[1] Johnson, M., "Advanced Capabilities for Emergency Response Operations (ACERO)," *NASA Technical Reports Server*, 2022. `https://ntrs.nasa.gov/citations/20220015931`.

[2] Wiggins, E. B., Anderson, B. E., Brown, M. D., Campuzano-Jost, P., Chen, G., Crawford, J., Crosbie, E. C., Dibb, J., DiGangi, J. P., Diskin, G. S., Fenn, M., Gallo, F., Gargulinski, E. M., Guo, H., Hair, J. W., Halliday, H. S., Ichoku, C., Jimenez, J. L., Jordan, C. E., Katich, J. M., Nowak, J. B., Perring, A. E., Robinson, C. E., Sanchez, K. J., Schueneman, M., Schwarz, J. P., Shingler, T. J., Shook, M. A., Soja, A. J., Stockwell, C. E., Thornhill, K. L., Travis, K. R., Warneke, C., Winstead, E. L., Ziemba, L. D., and Moore, R. H., "Reconciling Assumptions in Bottom-Up and Top-Down Approaches for Estimating Aerosol Emission Rates From Wildland Fires Using Observations From FIREX-AQ," *Journal of Geophysical Research: Atmospheres*, Vol. 126, No. 24, 2021, p. e2021JD035692. https://doi.org/10.1029/2021JD035692, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JD035692, e2021JD035692 2021JD035692.

[3] Wiggins, E. B., Soja, A. J., Gargulinski, E., Halliday, H. S., Pierce, R. B., Schmidt, C. C., Nowak, J. B., DiGangi, J. P., Diskin, G. S., Katich, J. M., Perring, A. E., Schwarz, J. P., Anderson, B. E., Chen, G., Crosbie, E. C., Jordan, C., Robinson, C. E., Sanchez, K. J., Shingler, T. J., Shook, M., Thornhill, K. L., Winstead, E. L., Ziemba, L. D., and Moore, R. H., "High Temporal Resolution Satellite Observations of Fire Radiative Power Reveal Link Between Fire Behavior and Aerosol and Gas Emissions," *Geophysical Research Letters*, Vol. 47, No. 23, 2020, p. e2020GL090707. https://doi.org/10.1029/2020GL090707, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL090707, e2020GL090707 2020GL090707.

[4] Ambrosia, V. G., Wegener, S., Zajkowski, T., Sullivan, D. V., Buechel, S., Enomoto, F., Lobitz, B., Johan, S., Brass, J., and Hinkley, E., "The Ikhana unmanned airborne system (UAS) western states fire imaging missions: from concept to reality (2006–2010)," *Geocarto International*, Vol. 26, No. 2, 2011, pp. 85–101. https://doi.org/10.1080/10106049.2010.539302, URL https://doi.org/10.1080/10106049.2010.539302.

[5] NIFC/USFS, "NIROPS Databse," https://ftp.wildfire.gov/public/nirops/for_testing, 2023. Accessed: October 20 2023.

[6] "Forest Service Autonomous Modular Scanner (AMS) Testing, Integration and Operational Use," USDA Forest Service poster, 2014. URL https://fsapps.nwcg.gov/nirops/files/poster-brochure-docs/AMS_Poster_Esize.pdf.

[7] USDA Forest Service, "Fire Imaging Technologies Users Guide," , 2020. URL https://www.nifc.gov/sites/default/files/document-media/Fire%20Imaging%20Technologies%20Users%20Guide.pdf.

[8] "Infrared Branch," National Interagency Fire Center (NIICD/NIROPS) web page, 2025. URL https://www.nifc.gov/resources/niicd/infrared-branch, accessed 2025-11-10.

[9] "Phoenix Capabilities and Products," NIROPS/NIICD web page, 2025. URL https://www.nifc.gov/resources/niicd/infrared-branch/phoenix-capabilities, accessed 2025-11-10.

[10] "FIRIS (Fire Integrated Real-time Intelligence System)," California Governor's Office of Emergency Services (Cal OES) web page, 2025. URL https://www.caloes.ca.gov/office-of-the-director/operations/response-operations/fire-rescue/firis/, accessed 2025-11-10.

[11] "WIFIRE: Fire Integrated Real-time Intelligence System (FIRIS)," UC San Diego WIFIRE Lab web page, 2025. URL https://wifire.ucsd.edu/firis-in-depth, accessed 2025-11-10.

[12] "Colorado Multi-Mission Aircraft (MMA) Briefing Paper," Tech. rep., Colorado Division of Fire Prevention and Control, 2017. URL https://ftp.wildfire.gov/public/incident_specific_data/calif_n/2017%20FEDERAL_Incidents/MMA_Support_2017/MMA%20Briefing%20Paper.pdf.

[13] "Colorado MMA and the Colorado Wildfire Information Management System (CO-WIMS)," Tech. rep., Colorado Division of Fire Prevention and Control, 2015. URL https://wildfiretoday.com/wp-content/uploads/2015/04/MMA-and-CO-WIMS.pdf.

[14] "NASA Uses New Technology to Understand California Wildfires," JPL News Release, February 28 2025. URL https://www.jpl.nasa.gov/news/nasa-uses-new-technology-to-understand-california-wildfires.

[15] "c-FIRST Team Sets Sights on Future Fire-observing Satellite Constellations," NASA Science—Technology Highlights, June 3 2025. URL https://science.nasa.gov/science-research/science-enabling-technology/technology-highlights/c-first-team-sets-sights-on-future-fire-observing-satellite-constellations/.

[16] Schroeder, W., Oliva, P., Giglio, L., Quayle, B., Lorenz, E., and Morelli, F., "Active fire detection using Landsat-8/OLI data," *Remote Sensing of Environment*, Vol. 185, 2016, pp. 210–220. https://doi.org/10.1016/j.rse.2015.08.032, URL https://www.sciencedirect.com/science/article/pii/S0034425715301206, landsat 8 Science Results.

[17] Murphy, S. W., de Souza Filho, C. R., Wright, R., Sabatino, G., and Correa Pabon, R., "HOTMAP: Global hot target detection at moderate spatial resolution," *Remote Sensing of Environment*, Vol. 177, 2016, pp. 78–88. https://doi.org/10.1016/j.rse.2016.02.027, URL https://www.sciencedirect.com/science/article/pii/S0034425716300554.

[18] Kumar, S. S., and Roy, D. P., "Global operational land imager Landsat-8 reflectance-based active fire detection algorithm," *International Journal of Digital Earth*, Vol. 11, No. 2, 2018, pp. 154–178. https://doi.org/10.1080/17538947.2017.1391341, URL https://doi.org/10.1080/17538947.2017.1391341.

[19] de Almeida Pereira, G. H., Fusioka, A. M., Nassu, B. T., and Minetto, R., "Active Fire Detection in Landsat-8 Imagery: a Large-Scale Dataset and a Deep-Learning Study," *CoRR*, Vol. abs/2101.03409, 2021. URL https://arxiv.org/abs/2101.03409.

[20] Zhan, J., Hu, Y., Cai, W., Zhou, G., and Li, L., "PDAM-STPNNet: A Small Target Detection Approach for Wildland Fire Smoke through Remote Sensing Images," *Symmetry*, Vol. 13, No. 12, 2021. https://doi.org/10.3390/sym13122260, URL https://www.mdpi.com/2073-8994/13/12/2260.

[21] Perrolas, G., Niknejad, M., Ribeiro, R., and Bernardino, A., "Scalable Fire and Smoke Segmentation from Aerial Images Using Convolutional Neural Networks and Quad-Tree Search," *Sensors*, Vol. 22, 2022, p. 1701. https://doi.org/10.3390/s22051701.

[22] Zhao, Y., Ma, J., Li, X., and Zhang, J., "Saliency Detection and Deep Learning-Based Wildfire Identification in UAV Imagery," *Sensors*, Vol. 18, 2018, p. 712. https://doi.org/10.3390/s18030712.

[23] Toulouse, T., Rossi, L., Campana, A., Celik, T., and Akhloufi, M. A., "Computer vision for wildfire research: An evolving image dataset for processing and analysis," *Fire Safety Journal*, Vol. 92, 2017, pp. 188–194. https://doi.org/10.1016/j.firesaf.2017.06.012, URL https://www.sciencedirect.com/science/article/pii/S0379711217302114.

[24] Foggia, P., Saggese, A., and Vento, M., "Real-Time Fire Detection for Video-Surveillance Applications Using a Combination of Experts Based on Color, Shape, and Motion," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 25, No. 9, 2015, pp. 1545–1556. https://doi.org/10.1109/TCSVT.2015.2392531.

[25] Kyrkou, C., and Theocharides, T., "Deep-Learning-Based Aerial Image Classification for Emergency Response Applications Using Unmanned Aerial Vehicles," *CVPR Workshops (UAVision)*, 2019. URL https://openaccess.thecvf.com/content_CVPRW_2019/papers/UAVision/Kyrkou_Deep-Learning-Based_Aerial_Image_Classification_for_Emergency_Response_Applications_Using_Unmanned_CVPRW_2019_paper.pdf.

[26] Kyrkou, C., and Theocharides, T., "AIDER (Aerial Image Dataset for Emergency Response Applications)," Zenodo, 2020. https://doi.org/10.5281/zenodo.3888300, URL https://zenodo.org/records/3888300.

[27] Shamsoshoara, A., Afghah, F., Razi, A., Zheng, L., Fulé, P. Z., and Blasch, E., "Aerial Imagery Pile Burn Detection Using Deep Learning: The FLAME Dataset," *Computer Networks*, Vol. 193, 2021, p. 108001. https://doi.org/10.1016/j.comnet.2021.108001, URL https://arxiv.org/abs/2012.14036.

[28] Chen, X., O'Neill, S., Richter, K., Hudak, A. T., Bright, B. C., Zajkowski, T., and Afghah, F., "Wildland Fire Detection and Monitoring Using a Drone-Collected RGB/IR Image Dataset," *IEEE Access*, Vol. 10, 2022, pp. 121305–121317. https://doi.org/10.1109/ACCESS.2022.3219150, URL https://www.fs.usda.gov/pnw/pubs/journals/pnw_2022_chen001.pdf.

[29] Wang, M., Yue, P., Jiang, L., Yu, D., Tuo, T., Li, J., et al., "An Open Flame and Smoke Detection Dataset for Deep Learning in Remote Sensing Based Fire Detection," *Geo-spatial Information Science*, Vol. 28, No. 2, 2024, pp. 511–528. https://doi.org/10.1080/10095020.2024.2347922, dataset DOI: 10.57760/sciencedb.j00104.00103.

[30] Tran, A., Tran, M., Marti, E., Cothren, J., Rainwater, C., Eksioglu, S., and Le, N., "Land8Fire: A Complete Study on Wildfire Segmentation Through Comprehensive Review, Human-Annotated Multispectral Dataset, and Extensive Benchmarking," *Remote Sensing*, Vol. 17, No. 16, 2025, p. 2776. https://doi.org/10.3390/rs17162776, URL https://www.mdpi.com/2072-4292/17/16/2776.

[31] Pesonen, J., Raita-Hakola, A.-M., Joutsalainen, J., Hakala, T., Akhtar, W., Koivumäki, N., Markelin, L., Suomalainen, J., de Oliveira, R. A., Pölönen, I., Honkavaara, E., et al., "Boreal Forest Fire: UAV-collected Wildfire Detection and Smoke Segmentation Dataset," *Scientific Data*, Vol. 12, 2025, p. 1419. https://doi.org/10.1038/s41597-025-05634-0, URL https://www.nature.com/articles/s41597-025-05634-0.

[32] Kularatne, W., Shayegan, K. S., Casado, C. Á., Rajala, J., Hänninen, T., López, M. B., and Nguyen, N. L., "FireMan-UAV-RGBT," , 2024. https://doi.org/10.5281/zenodo.13732947, URL https://doi.org/10.5281/zenodo.13732947.

[33] Hopkins, B., ONeill, L., Marinaccio, M., Rowell, E., Parsons, R., Flanary, S., Nazim, I., Seielstad, C., and Afghah, F., "FLAME 3 Dataset: Unleashing the Power of Radiometric Thermal UAV Imagery for Wildfire Management," , 2024. URL https://arxiv.org/abs/2412.02831, kaggle subset: https://www.kaggle.com/datasets/brycehopkins/flame-3-nadir-thermal-plot-subset.

[34] Mowla, M. N., Asadi, D., and Tekeoglu, K. N., "UAVS-FDDB: UAVs-based Forest Fire Detection Database," , 2024. https://doi.org/10.17632/5m98kvdkyt.2, URL https://data.mendeley.com/datasets/5m98kvdkyt/2.

[35] Mowla, M. N., Asadi, D., Tekeoglu, K. N., Masum, S., and Rabie, K., "UAVs-FFDB: A high-resolution dataset for advancing forest fire detection and monitoring using unmanned aerial vehicles (UAVs)," *Data in Brief*, Vol. 55, 2024, p. 110706. https://doi.org/10.1016/j.dib.2024.110706.

[36] Zhang, Y., Rui, X., and Song, W., "A UAV-Based Multi-Scenario RGB-Thermal Dataset and Fusion Model for Enhanced Forest Fire Detection," *Remote Sensing*, Vol. 17, No. 15, 2025, p. 2593. https://doi.org/10.3390/rs17152593, URL https://www.mdpi.com/2072-4292/17/15/2593.

[37] Gomez, C., "GOES16 Wildfires Smoke Plumes Dataset," Kaggle, 2025. URL https://www.kaggle.com/datasets/colvertgomez/goes16-wildfires-smoke-plumes-dataset, accessed 2025-11-10.

[38] Ülkü, I., and Akagündüz, E., "A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D images," *CoRR*, Vol. abs/1912.10230, 2019. URL http://arxiv.org/abs/1912.10230.

[39] Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A., "The Pascal Visual Object Classes (VOC) challenge," *International Journal of Computer Vision*, Vol. 88, No. 2, 2009, p. 303–338. https://doi.org/10.1007/s11263-009-0275-4.

[40] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L., "Microsoft Coco: Common Objects in Context," *Computer Vision – ECCV 2014*, 2014, p. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.

[41] Long, J., Shelhamer, E., and Darrell, T., "Fully Convolutional Networks for Semantic Segmentation," *CVPR*, 2015.

[42] Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J., "Pyramid Scene Parsing Network," *CVPR*, 2017.

[43] Yu, F., and Koltun, V., "Multi-Scale Context Aggregation by Dilated Convolutions," *ICLR*, 2016.

[44] Ronneberger, O., Fischer, P., and Brox, T., "U-Net: Convolutional Networks for Biomedical Image Segmentation," , 2015.

[45] Huang, B., Lu, K., Audeberr, N., Khalel, A., Tarabalka, Y., Malof, J., Boulch, A., Le Saux, B., Collins, L., Bradbury, K., Lefèvre, S., and El-Saban, M., "Large-Scale Semantic Classification: Outcome of the First Year of Inria Aerial Image Labeling Benchmark," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 6947–6950. https://doi.org/10.1109/IGARSS.2018.8518525.

[46] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H., "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," *ECCV*, 2018.

[47] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B., "The Cityscapes Dataset for Semantic Urban Scene Understanding," *CVPR*, 2016.

[48] Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P. H. S., and Zhang, L., "Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers," *CVPR*, 2021.

[49] Strudel, R., Garcia, R., Laptev, I., and Schmid, C., "Segmenter: Transformer for Semantic Segmentation," *ICCV*, 2021.

[50] Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., and Torralba, A., "Scene Parsing through ADE20K Dataset," *CVPR*, 2017.

[51] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P., "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," *NeurIPS*, 2021.

[52] Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., and Girdhar, R., "Mask2Former: Masked-attention Mask Transformer for Universal Image Segmentation," *CVPR*, 2022.

[53] He, K., Gkioxari, G., Dollár, P., and Girshick, R., "Mask R-CNN," *ICCV*, 2017.

[54] Akhloufi, M. A., Couturier, A., and Castro, N. A., "Unmanned Aerial Vehicles for Wildland Fires: Sensing, Perception, Cooperation and Assistance," *Drones*, Vol. 5, No. 1, 2021. https://doi.org/10.3390/drones5010015, URL https://www.mdpi.com/2504-446X/5/1/15.

[55] Roy, D., Wulder, M., Loveland, T., C.E., W., Allen, R., Anderson, M., Helder, D., Irons, J., Johnson, D., Kennedy, R., Scambos, T., Schaaf, C., Schott, J., Sheng, Y., Vermote, E., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Hipple, J., Hostert, P., Huntington, J., Justice, C., Kilic, A., Kovalskyy, V., Lee, Z., Lymburner, L., Masek, J., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R., and Zhu, Z., "Landsat-8: Science and product vision for terrestrial global change research," *Remote Sensing of Environment*, Vol. 145, 2014, pp. 154–172. https://doi.org/10.1016/j.rse.2014.02.001, URL https://www.sciencedirect.com/science/article/pii/S003442571400042X.

[56] Lee, W., Kim, S., Lee, Y.-T., Lee, H.-W., and Choi, M., "Deep neural networks for wild fire detection with unmanned aerial vehicle," *2017 IEEE International Conference on Consumer Electronics (ICCE)*, 2017, pp. 252–253. https://doi.org/10.1109/ICCE.2017.7889305.

[57] Park, M., Tran, D. Q., Lee, S., and Park, S., "Multilabel Image Classification with Deep Transfer Learning for Decision Support on Wildfire Response," *Remote Sensing*, Vol. 13, No. 19, 2021. https://doi.org/10.3390/rs13193985, URL https://www.mdpi.com/2072-4292/13/19/3985.

[58] Frizzi, S., Bouchouicha, M., Ginoux, J.-M., Moreau, E., and Sayadi, M., "Convolutional neural network for smoke and fire semantic segmentation," *IET Image Processing*, Vol. 15, No. 3, 2021, pp. 634–647. https://doi.org/10.1049/ipr2.12046, URL https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/ipr2.12046.

[59] Li, M., Zhang, Y., Mu, L., Xin, J., Yu, Z., Jiao, S., Liu, H., Xie, G., and Yingmin, Y., "A Real-time Fire Segmentation Method Based on A Deep Learning Approach," *IFAC-PapersOnLine*, Vol. 55, No. 6, 2022, pp. 145–150. https://doi.org/10.1016/j.ifacol.2022.07.120, URL https://www.sciencedirect.com/science/article/pii/S2405896322005055, 11th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2022.

[60] Behari, N., Holbrook, H. T., Garrett, P., Ippolito, C. A., and Dolph, C., *Contextual Segmentation of Fire Spotting Regions Through Satellite-Augmented Autonomous Modular Sensor Imagery*, ARC, 2021, Chap. 1, p. 1. https://doi.org/10.2514/6.2022-0495, URL https://arc.aiaa.org/doi/abs/10.2514/6.2022-0495.

[61] Iandola, F. N., Moskewicz, M. W., Ashraf, K., Han, S., Dally, W. J., and Keutzer, K., "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size," *CoRR*, Vol. abs/1602.07360, 2016. URL http://arxiv.org/abs/1602.07360.

[62] Jiao, Z., Zhang, Y., Xin, J., Mu, L., Yi, Y., Liu, H., and Liu, D., "A Deep Learning Based Forest Fire Detection Approach Using UAV and YOLOv3," *2019 1st International Conference on Industrial Artificial Intelligence (IAI)*, 2019, pp. 1–5. https://doi.org/10.1109/ICIAI.2019.8850815.

[63] Redmon, J., and Farhadi, A., "YOLOv3: An Incremental Improvement," , 2018.

[64] Zhao, L., Zhi, L., Zhao, C., and Zheng, W., "Fire-YOLO: A Small Target Object Detection Method for Fire Inspection," *Sustainability*, Vol. 14, No. 9, 2022. https://doi.org/10.3390/su14094930, URL https://www.mdpi.com/2071-1050/14/9/4930.

[65] Hossain, F. A., Zhang, Y. M., and Tonima, M. A., "Forest fire flame and smoke detection from UAV-captured images using fire-specific color features and multi-color space local binary pattern," *Journal of Unmanned Vehicle Systems*, Vol. 8, No. 4, 2020, pp. 285–309. https://doi.org/10.1139/juvs-2020-0009, URL https://doi.org/10.1139/juvs-2020-0009.

[66] Ghali, R., Akhloufi, M. A., Jmal, M., Souidene Mseddi, W., and Attia, R., "Wildfire Segmentation Using Deep Vision Transformers," *Remote Sensing*, Vol. 13, No. 17, 2021. https://doi.org/10.3390/rs13173527, URL https://www.mdpi.com/2072-4292/13/17/3527.

[67] Giglio, L., Schroeder, W., and Justice, C. O., "The Collection 6 MODIS active fire detection algorithm and fire products," *Remote Sensing of Environment*, Vol. 178, 2016, pp. 31–41. https://doi.org/10.1016/j.rse.2016.02.054.

[68] Schroeder, W., Oliva, P., Giglio, L., and Csiszar, I., "The New VIIRS 375 m active fire detection data product: Algorithm description and initial assessment," *Remote Sensing of Environment*, Vol. 143, 2014, pp. 85–96. https://doi.org/10.1016/j.rse.2013.12.008.

[69] Allison, R. S., Johnston, J. M., Craig, G., and Jennings, S., "Airborne Optical and Thermal Remote Sensing for Wildfire Detection and Monitoring: A Review," *Sensors*, Vol. 16, No. 8, 2016, p. 1310. https://doi.org/10.3390/s16081310.

[70] Ambrosia, V. G., Myers, J. S., and Hildum, E. A., "NASA's Autonomous Modular Scanner (AMS) – Wildfire Sensor: Improving Wildfire Observations From Airborne Platforms," *Proceedings of the 34th International Symposium on Remote Sensing of the Environment (ISRSE-34)*, ISPRS, Sydney, New South Wales, Australia, 2011, p. ?? Paper no. 472 (Ref. 472) in the programme.

[71] National Aeronautics and Space Administration (NASA), Airborne Sensor Facility, "AMS Flight 07-001-03, 28 Oct 2006, Esperanza Fire, Southern California (day/night)," https://asapdata.arc.nasa.gov/ams/flight_html/0700103.html, 2006. Data from AMS Land 2.5 (12-Channel) configuration; 29 tracks, 108 951 scanlines; principal investigator Ambrosia (CSUMB).

[72] Moeller, C., and Gumley, L., "MODIS Airborne Simulator Infrared Calibration," Tech. rep., Space Science and Engineering Center (SSEC), University of Wisconsin–Madison, Madison, WI, USA, 1999. URL https://bin.ssec.wisc.edu/pub/gumley/MAS/MAS_IR.pdf, technical report; unpublished internal SSEC memo.

[73] Hook, S. J., Myers, J. J., Thome, K. J., Fitzgerald, M., and Kahle, A. B., "The MODIS/ASTER airborne simulator (MASTER)—a new instrument for earth science studies," *Remote Sensing of Environment*, Vol. 76, No. 1, 2001, pp. 93–102. https://doi.org/10.1016/S0034-4257(00)00195-4, URL https://asapdata.arc.nasa.gov/sensors/master/reference/master.pdf.

[74] Kingma, D. P., and Ba, J., "Adam: A Method for Stochastic Optimization," *CoRR*, Vol. abs/1412.6980, 2014. URL https://api.semanticscholar.org/CorpusID:6628106.