remote sensing

MDPI

Article

Development and Application of Self-Supervised Machine Learning for Smoke Plume and Active Fire Identification from the Fire Influence on Regional to Global Environments and Air Quality Datasets

Nicholas LaHaye [1,2,*], Anastasija Easley [2,3], Kyongsik Yun [2], Hugo Lee [2], Erik Linstead [4,5], Michael J. Garay [2] and Olga V. Kalashnikova [2]

[1] Spatial Informatics Group, LLC., Pleasanton, CA 94566, USA
[2] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91101, USA; kyongsik.yun@jpl.nasa.gov (K.Y.); huikyo.lee@jpl.nasa.gov (H.L.); michael.j.garay@gmail.com (M.J.G.); olga.kalashnikova@jpl.nasa.gov (O.V.K.)
[3] Department of Mathematics, College of Letters & Science, University of California at Berkeley, Berkeley, CA 94720, USA
[4] Fowler School of Engineering, Chapman University, Orange, CA 92866, USA; linstead@chapman.edu
[5] Machine Learning and Assistive Technology Laboratory (MLAT), Chapman University, Orange, CA 92866, USA
* Correspondence: nlahaye@sig-gis.com

Abstract: Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ) was a field campaign aimed at better understanding the impact of wildfires and agricultural fires on air quality and climate. The FIREX-AQ campaign took place in August 2019 and involved two aircraft and multiple coordinated satellite observations. This study applied and evaluated a self-supervised machine learning (ML) method for the active fire and smoke plume identification and tracking in the satellite and sub-orbital remote sensing datasets collected during the campaign. Our unique methodology combines remote sensing observations with different spatial and spectral resolutions. With as much as a 10% increase in agreement between our produced masks and high-certainty hand-labeled pixels, relative to evaluated operational products, the demonstrated approach successfully differentiates active fire pixels and smoke plumes from background imagery. This enables the generation of a per-instrument smoke and active fire mask product, as well as smoke and fire masks created from the fusion of selected data from independent instruments. This ML approach has the potential to enhance operational active wildfire monitoring systems and improve decision-making in air quality management through fast smoke plume identification and tracking and could improve climate impact studies through fusion data from independent instruments.

Keywords: FIREX-AQ; active fire and smoke detection; big data applications; clustering; computer vision; image segmentation; self-supervised deep learning; multi-modal data fusion

Academic Editor: Xiaoyang Zhang

Received: 25 January 2025
Revised: 19 March 2025
Accepted: 24 March 2025
Published: 2 April 2025

Citation: LaHaye, N.; Easley, A.; Yun, K.; Lee, H.; Linstead, E.; Garay, M.J.; Kalashnikova, O.V. Development and Application of Self-Supervised Machine Learning for Smoke Plume and Active Fire Identification from the Fire Influence on Regional to Global Environments and Air Quality Datasets. *Remote Sens.* **2025**, *17*, 1267. https://doi.org/10.3390/rs17071267

1. Introduction

An important and common application of machine learning (ML) is to identify and leverage latent patterns in data or imagery. A typical approach is to use *supervised* learning, which requires a set of truth labels that the ML method attempts to generalize to the problem of mapping from an input dataset X to the output Y through a set of features, M. The challenge

with supervised learning, and even the recently popularized semi-supervised learning, is acquiring a sufficiently large and **unambiguous** set of labels, which often requires many hours of manual labor on the part of domain experts. Alternatively, *self-supervised* learning takes a similar input dataset X and finds relationships among the features M resulting in context-free groupings in the output Y. Because no labels are provided for the input, there are no labels provided in the output. To utilize the results, the labels or missing context must be assigned after the fact by experts, but this has proven to be a much less labor-intensive endeavor, all while keeping subject matter experts in the loop.

In previous work, we demonstrated that feeding 2-dimensional images of instrument radiances, or Level 1 (L1) data, into Deep Belief Networks (DBNs) coupled with an unsupervised clustering method results in images automatically segmented into relevant geophysical objects [1]. We further demonstrated that the same results can be achieved using a simplified architecture across select areas of the globe and for various kinds of land surface and atmospheric segmentation tasks [2].

In our recent work [3], we have generalized our ML framework into an open-source software system called Segmentation, Instance Tracking, and data Fusion Using multi-SEnsor imagery (SIT-FUSE). This framework allows for various types of encoders, including regular and convolutional DBNs, Transformers, and Convolutional Neural Networks (CNNs), and we have moved from traditional unsupervised clustering to a deep learning-based clustering approach.

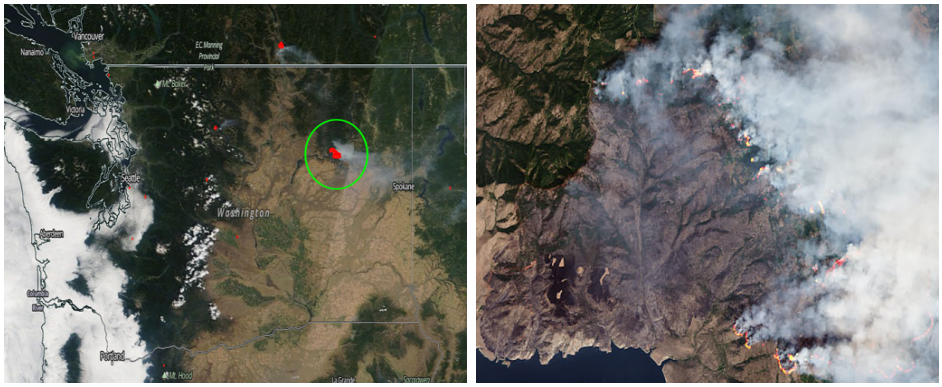
This approach, as a whole, has several unique benefits. First, it is not restricted to a particular remote sensing instrument with specific spatial or spectral resolution. Second, it has the potential to identify and "track" geophysical objects across datasets acquired from multiple instruments. Third, it allows for the joining of data from different instruments, "fusing" the information within the self-supervised encoder. Finally, it can be applied to many different scenes and problem sets, most notably in no- and low-label environments, not just ones for which labeled training sets exist, which is required for strictly supervised ML techniques.

Here, we apply our self-supervised ML approach to the problem of automatically detecting and tracking active wildfires and smoke plumes, through sequences of open-access L1 (imagery) data acquired by multiple remote sensing instruments during the joint National Aeronautics and Space Administration/National Oceanic and Atmospheric Administration (NASA/NOAA) Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ) field campaign that took place in the western United States in the summer of 2019 [4]. The high-altitude NASA ER-2 carried seven remote sensing instruments that provided high-spatial-resolution observations of active fires and smoke plumes in conjunction with NASA DC-8 aircraft and multiple satellite overpasses over the same fire events. The FIREX-AQ datasets of collocated satellites and multiple airborne imagery at different spatial resolutions are excellent as a testbed for the SIT-FUSE-based method of active fire/smoke identification and tracking, for which we have released the intermediate and final outputs for public access.

Wildfires and the smoke plumes induced by wildfires substantially contribute to the carbon cycle and can have a long-lasting impact on air quality and Earth's climate system. In addition, human-driven climate change is associated with more frequent and severe wildfires [5]. Despite the importance and immediacy of the problem, most research and decision-support tools to study wildfires and plumes use observations from a single instrument whose spatial coverage and (spatial, spectral, and temporal) resolutions vary from very fine to very coarse scales, neither of which, on their own, is fully capable of providing the much-needed information for a comprehensive understanding of wildfires and wildfire smoke [6]. As such, the current study aims to combine datasets with different spatial resolutions from multiple instruments to create a patchwork of datasets that fill the

temporal gaps present in current single-instrument active fire detection datasets. Here, the first step is testing a general framework for segmenting the datasets from multiple instruments and identifying wildfires and smoke plumes. Figure 1 shows a map of the active fire (red area within the green circle), taken from NASA's WorldView Snapshots web tool and a close-up reference image of the Williams Flats fire, one of the fires we focus on within this study, taken from the Landsat-8 Operational Land Imager (OLI).

The detection and tracking of objects, like wildfires and smoke plumes, within a single-instrument dataset has long required developing instrument-specific retrieval algorithms. Such development is labor-intensive and requires domain-specific parameters and instrument-specific calibration metrics, alongside the manual effort to track retrieved objects across multiple scenes [7]. The recent development of retrieval algorithms is actively underway in the field of supervised deep learning (DL), and various methods (e.g., Convolutional Neural Networks (CNNs)) have been applied. Some of these DL methodologies work well, in terms of precision and accuracy [7], but are still limited by the requirement that the spatial resolutions between training datasets and output products be the same. These methods also require pre-existing label sets, unlike recent supervised approaches like Fully Convolutional Networks (FCNs), Mask R-CNNs, and Transformers [8–10], which require large label sets to archive accurate results.



(**a**) (**b**)

Figure 1. Reference map and imagery. (**a**) Map of fire location from publicly available NASA WorldView Snapshots/MODIS [11,12]. (**b**) Publicly available reference Williams Flats fire image from Landsat-8/OLI [13].

In our previous work, we demonstrated that an encoder trained in a self-supervised manner, namely a Deep Belief Network (DBN), trained with L1 (instrument radiance) images, can segment images based on geophysical objects within the scene, in conjunction with unsupervised clustering [1]. The unique benefit of this method is that its application is not limited to a single spatial or spectral resolution, and the method has the potential to detect and track objects from images with different resolutions from multiple instruments. With this method, instead of requiring a per-instrument finely hand-labeled label set, we can apply a coarser manual context assignment after segmentation on a smaller set of training scenes, allowing for this technique to be easily applied in cases of no labels or limited labels. We have also quantitatively validated that the same could be achieved using a simpler architecture for a set of atmospheric and land surface classification tasks using varying spectral, spatial, temporal, and multi-angle remote sensing data as input [2]. Since this work, we have transitioned from unsupervised clustering to self-supervised deep clustering, which we will discuss further in the Section 2.3. This completely self-supervised approach can leverage training data from many different scenes, not just ones that are accounted for by previous label sets for training, as is the case with strictly supervised

techniques. Ongoing research applies this self-supervised machine learning methodology to track detected smoke plumes across spatiotemporal domains. However, this study focuses on identifying wildfire and smoke plumes within a single-instrument dataset and using a fusion of datasets from multiple instruments.

This approach not only allows us to leverage single- and multi-instrument datasets to create a denser static patchwork of active fire and smoke detections with increased spatial, spectral, and temporal resolution (as depicted in Figures 2–4), but it also gives us a uniform embedding-based representation of the data via the encoder outputs and final output of clusters. The final cluster output can be used in conjunction with spatial distributions of the output labels to facilitate active fire and smoke plume instance tracking across multi-sensor scenes over varying spatiotemporal domains. Figures 2–4 demonstrate the various tiers and scales of representative capabilities over the Williams Flats fire on 6 August 2019, when incorporating observations GOES at the coarse spatial but fine temporal resolution end of the scale, and the airborne instruments mentioned in Table 1 over the Williams Flats and Sheridan fires at the fine spatial but coarse temporal end of the scale, along with the polar orbiters in-between these two extremes.

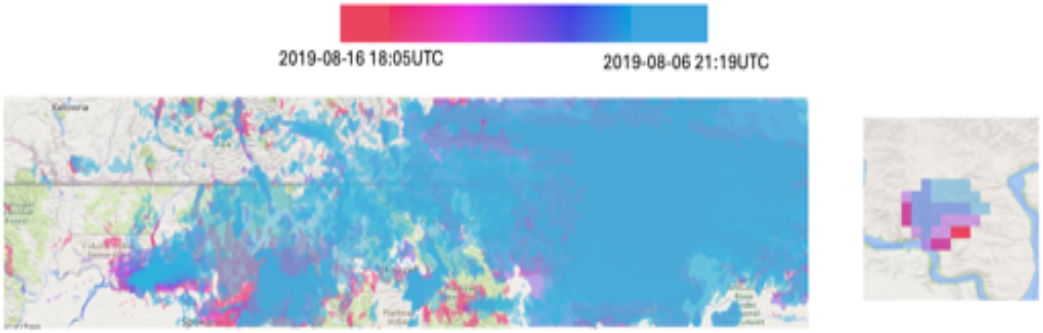


Figure 2. A display of the smoke plume (**left**) and the active fire front (**right**) progression of the Williams Flats fire on 6 August 2019, as captured by GOES-17 and segmented via SIT-FUSE.

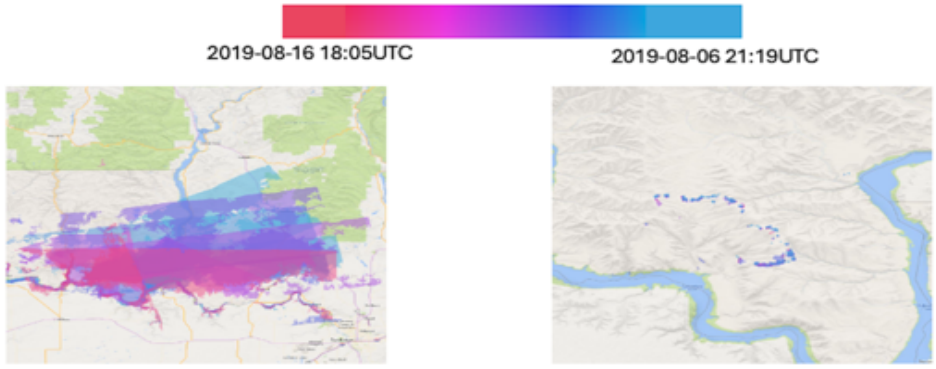


Figure 3. A display of the smoke plume (**left**) and the active fire front (**right**) progression of the Williams Flats fire on 6 August 2019, as captured by eMAS and segmented via SIT-FUSE. When used in conjunction with the data in Figure 2 we can look across large spatial domains at fine temporal scales as well as fine-scale detail at higher spectral resolutions.

Work on the general problem of self-supervised image segmentation appears to have had success in separating the foreground from the background [14,15], or have made significant subsets of spectral resolution (using a single band of input) from one type of instrumentation, which is effective for their applications, but does not provide the spectral specificity or per-observation or temporal resolution we are trying to attain here [16]. Other works have focused on urban planning and mapping, outlining buildings and roadways [17], which is not the goal here. A similar study that used a similar machine learning

approach to us—using autoencoders for representation learning and clustering for unsupervised segmentation—attained an accuracy of 83% on Landsat imagery alone [18]. This uses a similar kind of model to our studies but uses a single instrument. With large variations in spatial and spectral resolutions, our technique attains higher accuracy (and balanced accuracy, in some cases) across many different instrument sets, including fused data. Even with more recent breakthroughs in semi-supervised semantic segmentation, like the Segment Anything Model (SAM), a problem-dependent amount of labels is required, and SAM is largely unproven in complex domains like remote sensing [19]. The identification of the necessary size of label sets, generation of per-pixel label sets, and testing of the feasibility of new techniques in more complex domains are all problem-specific and time-consuming tasks that can be skipped, given our solution—as seen in the successful but extremely limited cases discussed in [20–23]. Lastly, there are new physics-based retrieval techniques, which seem promising, but need continued rigorous analysis to generalize across different regions and instrument types [24]. In the future, it may be useful to combine the physical parameterizations and ML-based retrievals via ML loss functions that are "physics-aware". The lack of need for large new label sets mitigates the costly, labor-intensive work of manually segmenting each pixel within a dataset used for ground truth, a process which is itself error-prone, and other previously mentioned supervised learning-related precursors model training. Also, leveraging pre-existing operational products to use as labels for supervised learning tasks will inherently cause them to either lack training set diversity or suffer from the issues mentioned above. On the other hand, our approach is well suited to handle large amounts of data, because our unsupervised and self-supervised models can perform label-free image segmentation. The fact that the human-in-the-loop steps of context application and validation occur after the images have been segmented allows for human oversight while mitigating the need for the extremely labor-intensive act of pixel-by-pixel manual segmentation for tens of thousands of images. In the subsequent sections, we will describe the experimental design for evaluating the performance and efficacy of using SIT-FUSE in support of Fire Influence on Regional to Global Environments Experiment—Air Quality 2019 (FIREX-AQ 2019; https://csl.noaa.gov/projects/firex-aq/, accessed on 20 December 2024), the results, the conclusion of the experiments, provide further discussion points, and discuss current and future work on this approach, the associated framework, and the correlated tooling.

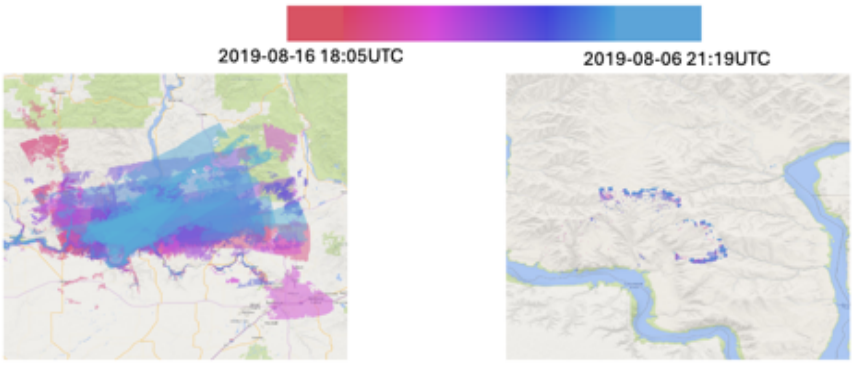


Figure 4. A display of the smoke plume (**left**) and the active fire front (**right**) progression of the Williams Flats fire on 6 August 2019, as captured by eMAS, MASTER, AirMSPI, and AVIRIS-C and segmented via SIT-FUSE. This combination of instrumentation maximizes the temporal resolution at the associated high spatial and spectral (/polarization) resolutions of the airborne instrumentation. When compared to the data in Figure 3, this increases the yields even more for monitoring and science capabilities when used in conjunction with geospatial and polar orbiting instrumentation.

Table 1. Airborne instruments and their products.

Platform	Instruments	Science Products	Spatial Resolution
NASA ER-2	Airborne Multiangle SpectroPolarimetric Imager (AirMSPI) [25]	Spectro-polarimetric intensities (10 m spatial resolution, 8 wavelengths in 355–935 nm spectral range, 3 polarimetric bands)	10 m
NASA ER-2	Enhanced MODIS Airborne Simulator (eMAS) [26]	Spectral intensities in 38 bands in 445–967 nm and 1.616–14.062 µm spectral ranges	50 m
NASA DC-8	MODIS/ASTER Airborne Simulator (MASTER) [27]	Spectral intensities in 50 bands in 0.44–12.6 µm spectral range	10–30 m
NASA DC-8	Airborne Visible/Infrared Imaging Spectrometer—Classic (AVIRIS-C) [28]	Spectral intensities in 224 bands in 400–2500 nm spectral range	10–30 m

2. Materials and Methods

2.1. Input Data

To demonstrate the efficacy of the approach for multi-platform smoke and active fire front segmentation for field campaign support, we used remote sensing datasets from airborne and satellite instruments collected during the FIREX-AQ 2019 campaign. Specifically, we investigated the wildfires and smoke plumes originating from the Williams Flats fire, Horsefly fire, and Mosquito fire over the three days from 6 August through 8 August 2020. Where available, we also tested on scenes over the Sheridan Fire, from 15 August through 23 August. The combination of different fires and associated landscapes for the central fires of the campaign (Williams Flats and Sheridan) as well as two additional fires (Horsefly and Mosquito) give us geospatial, land cover, and temporal variance to evaluate over while demonstrating efficacy in the field campaign setting. The FIREX-AQ campaign involved two aircraft: the DC-8 aircraft with a primary payload of in situ instruments supplemented by several remote sensors, and NASA's high-altitude ER-2 aircraft with a package of seven remote sensing instruments. The ER-2 aircraft coordinated on a few occasions with the DC-8 aircraft and had multiple collocations with satellite sensors flying directly along the satellite track. Therefore, several observations of the same active fires and smoke plumes were made by multiple instruments at various spatial resolutions. Tables 1 and 2 summarize the datasets from airborne and satellite instruments used in this study.

Table 2. Satellite instruments and their products.

Platform	Instruments	Science Products	Spatial Resolution
Terra	Multi-angle Imaging SpectroRadiometer (MISR) [29]	Spectral intensities at 446 nm, 558 nm, 672 nm, and 867 nm	1.1 km and 275 m, all resampled to 1.1 km
Terra and Aqua	MODerate resolution Imaging SpectroRadiometer (MODIS) [11]	Spectral intensities in 38 bands in 445 nm–967 nm and 1.616 µm—14.062 µm spectral range	1 km when used alone and resampled to 1.1 km when used with MISR
Suomi NPP and NOAA-20	Visible Infrared Imaging Suite (VIIRS) [30]	Spectral intensities in 5 bands in 0.44–12.6 µm spectral range	375 m
GOES-17	Advanced Baseline Imager (ABI) [31]	Spectral intensities in 50 bands in 0.47–13.3 µm spectral range	1 km and 2 km, all resampled to 2 km
PlanetScope	Dove Imagers [32]	Spectral intensities in 4 bands in 455–860 nm spectral range	1–3 m native resolution, all resampled to 10 m per the NASA Technical Report https://ntrs.nasa.gov/citations/20240001694, accessed on 10 January 2025

2.2. Data Preprocessing

For single-instrument cases, data were re-projected to the WGS84 Latitude/Longitude projection. For multi-sensor cases, data are collocated, re-projected to the WGS84 Latitude/Longitude projection, resampled to the lowest common spatial resolution, and stacked

channel-wise. The actual fusion occurs as a part of the representation learning performed inside the encoder.

Samples were generated by taking pixels and their direct neighbors (center pixel + 8 neighbors), in all channels, and creating a flat vector. These samples were then standardized by removing the per-channel mean and scaling to per-channel unit variance. The per-channel mean and standard deviations are computed over the full sample set from the scenes used for training.

For training, the scenes used are subsets. To ensure the subset of samples from the training scenes contained a representative variety of terrain and phenomena, k-means clustering was applied to the data with a set of 50 classes. This stratification technique was chosen as it has proven to be an effective, albeit naive, way to ensure variation within training samples [33–35]. Concerning the number of classes chosen (50), we ran a simple elbow criteria analysis, identifying the number of classes that maps to the beginning of diminishing returns relative to the inertia, or the sum of squared distances between each point and the centroid of the cluster it has been assigned to. Figure 5 depicts the cluster count vs. inertia plots for each instrument. Three million samples are randomly sampled based on the stratification generated by the full number of samples from the training scenes labeled with the 50 k-means classes. All pixels that are set to fill values, or out of specified valid ranges are discarded before any preprocessing. Regarding the spectral bands used, all spectral bands were used except bands that were extremely noisy or known to be non-functional for the FIREX-AQ campaign time frame. The same pixels used to train the encoders are also used to train the deep clustering heads, and the full training scenes are used to assign context.

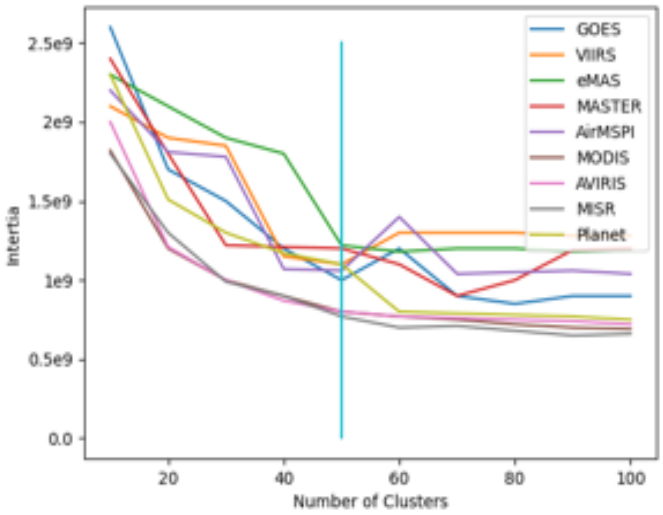


Figure 5. A set of the cluster count vs. inertia plots for each instrument that informed our selection of the number of clusters to use for the k-means-based naive data stratification approach we employ. The horizontal line is at K = 50, the selected number of clusters.

SIT-FUSE can take larger tiles for Convolutional DBNs, CNN-based, and Transformer-based architectures (as mentioned below). However, the pixel neighborhood has proven to provide enough spatial context with regular DBNs, as used in this work. Figure 6 is a flow chart that depicts the overall flow of SIT-FUSE.

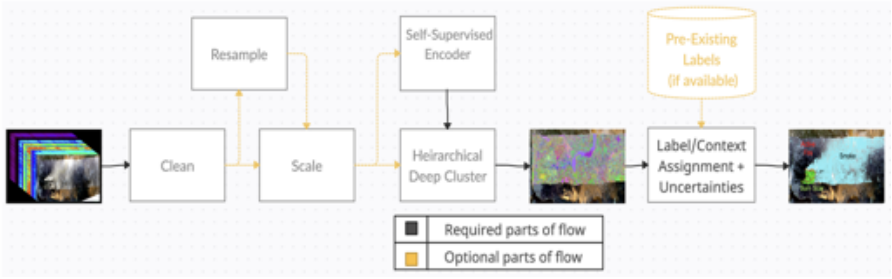


Figure 6. A flow diagram for the processing of one input type (single-instrument or fusion set) through SIT-FUSE.

2.3. Methods

2.3.1. Self-Supervised Representation Learning

SIT-FUSE is developed to be a generic framework allowing various kinds of encoders and foundation models that leverage self-supervised representation learning, including Deep Belief Networks (DBNs) trained using contrastive divergence, Convolutional Neural Networks (CNNs) with residual connections trained via Bootstrap Your Own Latent (BYOL), and Transformers trained using Image-Joint Embedding Predictive Architecture (I-JEPA) or Masked AutoEncoders (MAEs), as well as pre-trained Transformers for Earth Science, like Clay (https://madewithclay.org/, accessed on 10 November 2024) [36–39]. For all of these experiments, we used DBNs with 2–3 layers. DBNs were selected here because previous work and experiments performed in this work demonstrated that they produce reasonable results and the parameter space is much smaller than the other models mentioned above. We have performed extensive validation on the use of DBNs from both the perspective of structural understanding and downstream task performance, as well as resource consumption assessment, for the large set of single-instrument and fusion datasets [2]. In short, 2–3 layer DBNs provide a relatively compact model (2 million parameters vs. 100 million–10 billion parameters) with representational capabilities that meet our needs. While given the per-layer training paradigm and the generative nature that has caused many to turn to other architectures/approaches, recent work has demonstrated out-sized representational capabilities relative to other much larger generative models, like Generative Adversarial Networks (GANs) [38]. Given our successes and studies like [38], we continue to provide DBNs as an encoder option within SIT-FUSE, while also acknowledging the difficulties others have had with them in the past. We are currently evaluating encoder complexity in relation to segmentation performance and geographic coverage, to optimally operationalize this approach for operational global production, which will be discussed further in later manuscripts.

As in our previous work [1–3], the DBN architectures used leverage feature expansion, outputting embeddings of a pixel neighborhood in a larger feature space than the one with which the samples were input. Similar to the idea behind input kernelization, architecture-based feature expansion allows the models to learn nonlinear latent patterns that would be compact and complex in lower dimensions in simpler but higher-dimension forms. This method, although not the most common use case for DBNs or encoders in general, has been demonstrated effectively in other studies as well as in our previous work [1,2,40–42]. In previous works, we held architecture parameters static for all tests to demonstrate efficacy. Here, we varied the number of layers and hidden/output parameters for each encoder used, based on input data spectral resolution.

2.3.2. Deep Clustering

To extract segmentation maps from the per-pixel embeddings, we use deep clustering, specifically Information Invariant Clustering.

Previous experiments used BIRCH and other forms of traditional agglomerative clustering. We have transitioned to DL-based clustering because the training time, inference time, memory requirements, and model re-usability are all much improved when using neural network layers via PyTorch2.0 when compared to using traditional clustering via sci-kit-learn. These layers are trained using the invariant information clustering (IIC) loss function. IIC aims to assign labels that maximize mutual information between an input sample x and a perturbed version of x, x' [43]. For our use, perturbations are additions of Gaussian noise to the outputs of RBM-based architectures.

To mimic the hierarchical nature of agglomerative clustering output, we have set up hierarchical deep clustering layers. Here, the output heads are set up in a tree structure where each sub-tree is only trained on label samples classified as belonging to their parent label sets. Each layer receives only the output of the encoder, but as the tree is built from the top down, each neural network in a child node position on the tree only sees samples assigned labels associated with its parent and ancestor node(s). In this way, we can create scene segmentation at varying levels of specificity and explore the connections between each level. To our knowledge, this is the first study/software system that leverages IIC layers, or even deep clustering layers, in this fashion. For each of the pipelines, the root IIC head allows for 800 possible classes and each child IIC head allows for 100 subclasses. Currently, the level of hierarchy is specified by the user and there is no automated node splitting, but this could be implemented in the future. For our purposes, only two levels of hierarchy are used. We believe that this not only allows us to attain the specificity required for our segmentation tasks but that the hierarchical labeling can be leveraged as a co-pilot for data exploration and discovery. Figure 7 shows a two-level tree/flow diagram for performing deep clustering.

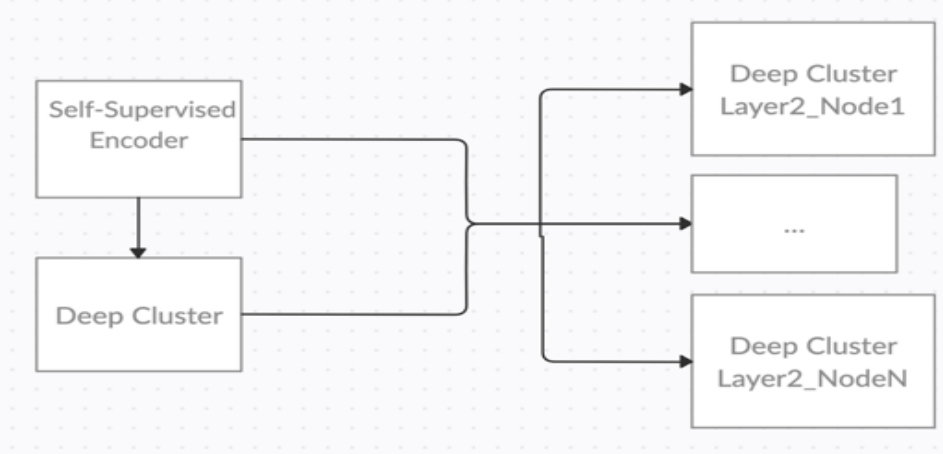


Figure 7. A 2-layer flow diagram example of the setup for hierarchical deep clustering. Each box labeled 'Cluster' is a set of fully connected layers, connected to the RBM-based model and trained via the IIC loss function. Each child node is only trained and makes predictions on samples given the label from its parent nodes that are associated with the path of edges that link the root to the current node. This setup allows us to use deep clustering to create interlaced levels of specificity for data exploration and characterization.

2.3.3. Hand Labeling

For both context assignment and validation, subject matter experts labeled areas of high-certainty smoke, active fire, and associated backgrounds. All areas that labelers were uncertain about, remained without labels. This labeling process was performed by generating polygons over the remote sensing imagery. An example is provided below in Figure 8. Because scenes can have overlapping classes (i.e., active fire and smoke are contained in the same pixel), but also have areas distinct to a single class, a separate

background class label set was generated for active fire and smoke. Figure 8 illustrates the labeling process over a single AVIRIS-C scene.

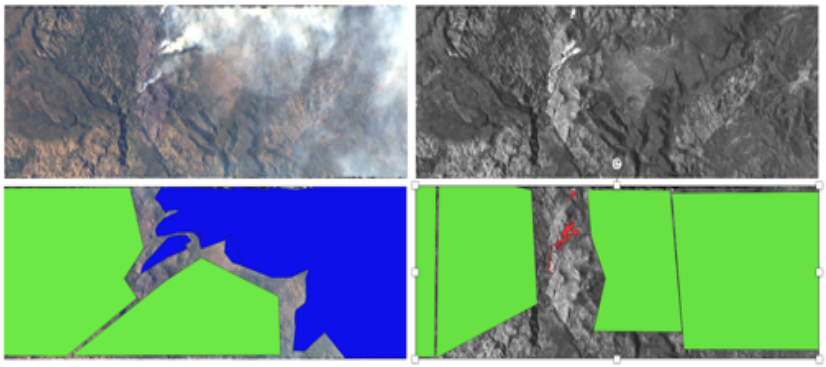


Figure 8. Visualization of the labeling process. The top row depicts a reference RGB generated from an AVIRIS-C scene (**top left**) and a single thermal band used for identifying active fire front locations (**top right**). The corresponding image in the bottom row has the smoke and smoke background labels over the RGB image and the active fire and fire background labels over the thermal image.

Unlike labeling for supervised learning, this approach does not require all training samples to be labeled, which is relevant for problems with high uncertainty in boundary cases, like the segmentation of active fire fronts and smoke plumes. The labeling of only the high-certainty class areas allows us to capture and compare against segmentation structure, and provide ample samples for context assignment. Although this labeling minimizes the pre-analysis labor required from subject matter experts, it still keeps experts in the loop (a crucial piece for science-related ML tasks). Relative to the number of scenes labeled here, supervised and semi-supervised tasks require 100× more labeled samples or more. Also, because they are learning the mapping between the labels and the input datasets directly, they require much more complete label sets (i.e., uncertain areas must be labeled background or foreground, potentially leading to systematic over-segmentation or under-segmentation).

2.3.4. Context Assignment

To assign context to the context-free segmentation maps generated via SIT-FUSE, we used zonal histogramming. For this, we overlayed the labels for a specific scene on the segmentation map generated for the same scene and generated counts of overlapping occurrences of the segmentation map classes and the high-certainty smoke, active fire, and background labels. These were gathered over multiple scenes, and then each class in the segmentation maps was assigned to the labels it best matched. This assignment was performed separately for each class of interest and its associated background class to ensure multi-class representation, like pixels containing both smoke and active fire. Figure 9 provides an example of labels overlayed on a segmentation map.

2.3.5. Contouring and Filtering

As the pixels within the detection are components of larger, often connected, geophysical objects, SIT-FUSE allows for building objects from separate detection pixels. To achieve this, we leverage the contouring capabilities within the 2.0 openCV Python/C++ software package [44,45]. Once generated, these contours are filled and we are left with multiple separate objects, instead of many more distinct pixels. We can also filter out small and large objects, where appropriate, and while not used here, leveraging dilations and erosions on the contoured objects is also possible. To mimic the automated process of an opera-

tionalized version and ensure fair comparisons and validation, all processing is performed uniformly across all scenes in an input set.

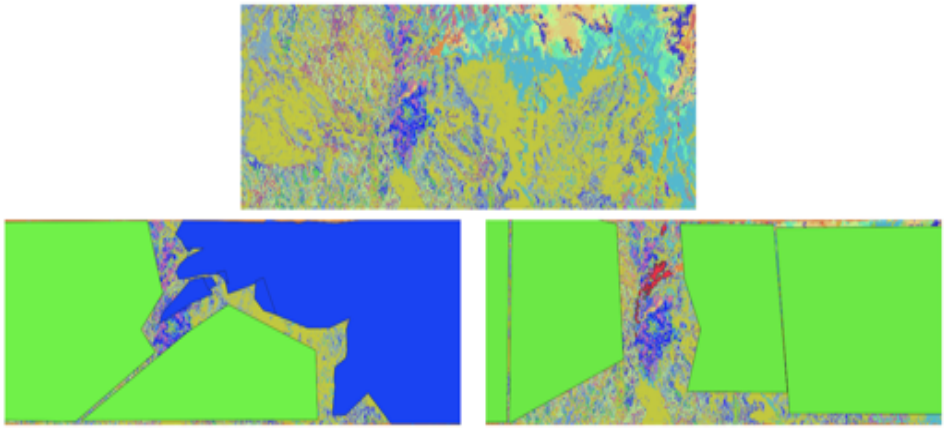


Figure 9. Visualization of the overlay required for the zonal histogramming process. The top row contains the context-free segmentation map generated via SIT-FUSE for the scene in Figure 8. The bottom row contains the same map with the smoke/background (**left**) and active fire/background (**right**) labels overlayed. The zonal histogramming process provides counts of classes in the segmentation map relative to label positions and each class is assigned to the label it best agrees with over multiple scenes.

2.3.6. Validation

There is no direct ground truth here—meaning comparisons are performed against pre-existing detections that are both used operationally and are considered the operational state of the art, as well as hand labels generated over only the high-certainty areas of smoke plumes and active fire fronts—and we know there are other areas of these objects that are not included in the labels. We include both comparisons as part of the validation because neither dataset is actual ground truth, and both comparisons provide valuable perspectives on performance. For each type, we split out the experiments and evaluation metrics to provide a clear picture of performance relative to the two different kinds of ground-truth-proxies used for performance evaluation in this study. Given the fact that there are areas known to be omitted from both datasets we are comparing against, and areas of false positives in the pre-existing detections, recall, precision, and their collective summary via F1-scoring are too harsh of evaluators here. We evaluated some previously published versions of precision and recall that apply fuzzy logic, but ultimately landed on the structural similarity index (SSIM) to evaluate performance across the various dataset pairs [46,47]. This is a fairly common problem within the remote sensing domain and one we aim to help solve with the collective incorporation of self-supervised learning, subject matter expert domain knowledge, and large amounts of data [48,49].

2.4. Materials and Tools

The software was developed with Python 3.9.13. SIT-FUSE has open-source functionality at its core [50]. To achieve the required goals of the software and leverage pre-existing and well-validated open-source software, geospatial, big data, and ML toolkits are the backbone of SIT-FUSE. For optimized handling and computation on large datasets across CPUs and GPUs, numpy, scipy, dask, xarray, Zarr, numba, and cupy are used [51–57]. For CPU- and GPU-based ML model training, deployment, evaluation, and auto-differentiation, sci-kit-learn, PyTorch, and torchvision are used [58,59]. Because RBMs are not included within the PyTorch library, Learnergy, an open-source library that contains various PyTorch-backed RBM-based architectures is used as well [60]. On the geospatial side of the problems being solved,

pyresample, GDAL, OSR, healpy, polar2grid, and GeoPandas are leveraged [61–65]. Lastly, for non-machine learning computer vision techniques, OpenCV is used [66]. The combination of these commonly used and well-tested software systems allows us to employ state-of-the-art approaches and architectures with minimal development and maintenance efforts, most of which are only minimally visible to the end user. SIT-FUSE is also publicly available and maintained on the public version of GitHub. For labeling, context assignment, and visualization/qualitative assessments, QGIS2.2, an open-source Geographic Information System (GIS) was used [67].

The hardware utilized was an NVIDIA GeForce Titan V100 GPU with 32 GB memory, as well as the NCCS Prism GPU Cluster (https://www.nccs.nasa.gov/systems/ADAPT/Prism, accessed on 7 October 2024).

3. Results

3.1. Fire Detection

Table 3 summarizes the performance of our SIT-FUSE-based segmentation approach, compared against the hand-labeled high-certainty active wildfire front areas, for active wildfire fronts across all scenes in the test set, for each instrument tested. Table 4 presents the same, but only for active fires and geographic areas not seen during training. In all cases, our approach performs well. When visually compared, our approach tends to over-segment relative to the high-certainty active fire pixels. However, the vast majority of the over-segmentation is over areas burning at lower heat, which are therefore less visibly associated with the active fire front. This effect can be seen for the Williams Flats fire in Figure 10 and for the Sheridan fire in Figure 11.

Table 3. Summary of active fire detection comparisons against the hand-labeled high-certainty active wildland fire fronts over the full set of test scenes that contain active fires. Total pixel count is the total number of pixels tested. SSIM was the metric used for comparisons in this study. For VIIRS and MODIS, the data from both platforms are used collectively.

Dataset	Total Pixel Count	SSIM
MASTER	12,438,023	0.73
eMAS	162,644,645	0.87
AVIRIS-C	177,233,951	0.89
GOES-17	1,016,785	0.82
VIIRS	4,759,386	0.88
MODIS	3,262,896	0.88

Table 4. Summary of active fire detection comparisons against the hand-labeled high-certainty active wildland fire fronts over only the scenes that contain the Sheridan fire and Horsefly fire, unseen during training. Total pixel count is the total number of pixels tested. SSIM was the metric used for comparisons in this study. For VIIRS and MODIS, the data from both platforms are used collectively.

Dataset	Total Pixel Count	SSIM
MASTER	1,708,746	0.89
eMAS	133,762,591	0.88
AVIRIS-C	21,073,834	0.93
GOES-17	2,541,975	0.84
VIIRS	1,189,846	0.88
MODIS	815,724	0.88

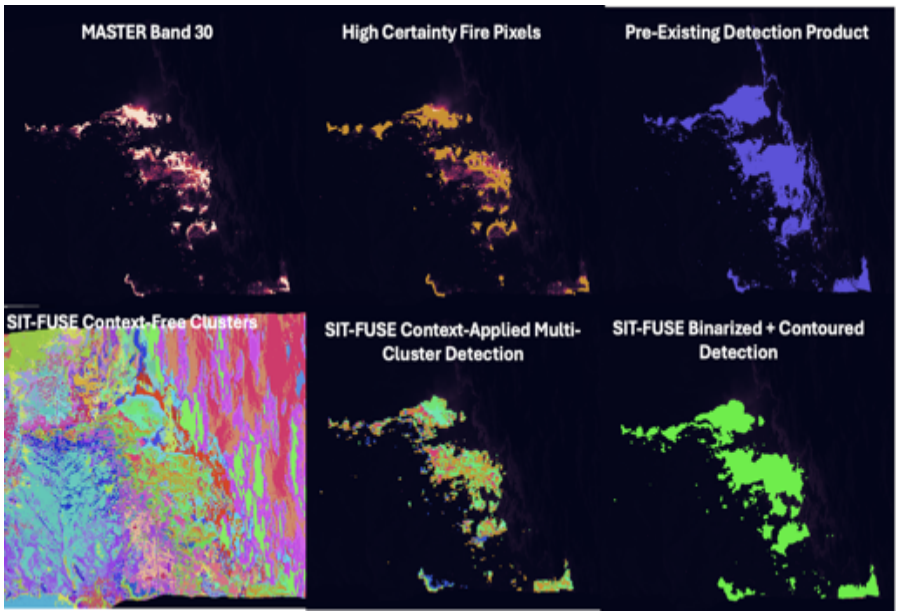


Figure 10. The top left panel is a single pseudo-colored thermal band from MASTER from a scene over the Williams Flats fire. Following this panel, on the top row are the hand-labeled high-certainty active fire pixels overlayed on the MASTER band, and the pre-existing band-ratio-based detection overlayed on the MASTER band, respectively. The bottom row consists of the context-free segmentation map generated from SIT-FUSE for this scene; the context-applied multi-cluster detection, which is a subset of the full segmentation map with only the labels/clusters that correlate to active fire fronts; and the final binarized and contoured active fire mask.

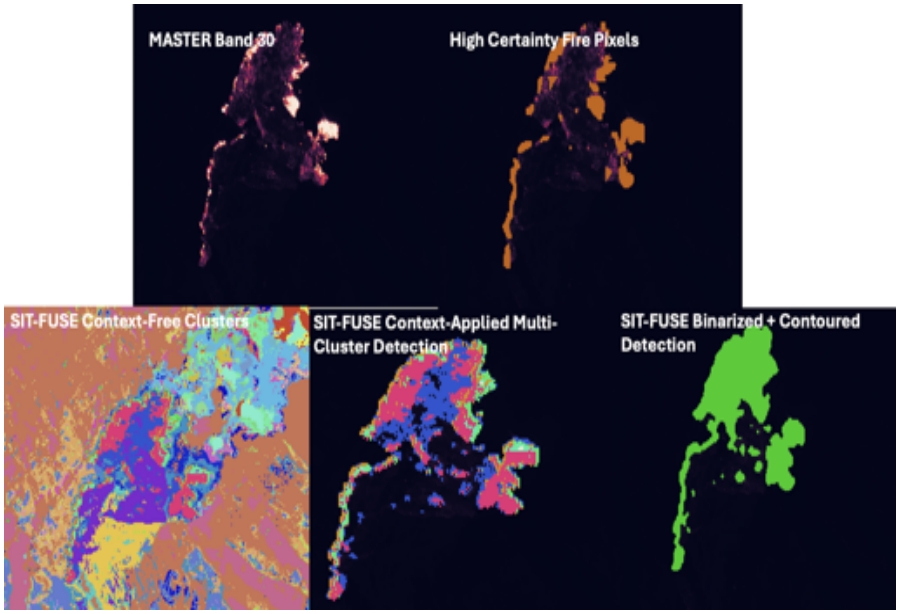


Figure 11. The top left panel is a single pseudo-colored thermal band from MASTER from a scene over the Sheridan fire. Following this panel, on the top row are the hand-labeled high-certainty active fire pixels overlayed on the MASTER band, and the pre-existing band-ratio-based detection overlayed on the MASTER band, respectively. The bottom row consists of the context-free segmentation map generated from SIT-FUSE for this scene; the context-applied multi-cluster detection, which is a subset of the full segmentation map with only the labels/clusters that correlate to active fire fronts; and the final binarized and contoured active fire mask.

Table 5 summarizes the same type of comparisons, this time against SIT-FUSE-based detections, and results from an experimental instrument-specific band-ratio-based detection methodology are compared against the high-certainty active fire pixel label set for MASTER.

For this comparison, we used all scenes in the test set with an associated band-ratio-based detection to compare against. The segmentation quality is quite comparable, with both techniques obtaining the same SSIM score; however, the difference maps in Figure 12 show a test-set-wide trend of hot smoke also being picked up as part of the active fire front in the band-ratio-based detection.

Table 5. Summary of active fire detection comparisons against the hand-labeled high-certainty active wildland fire fronts over only the scenes that contain the Sheridan fire and Horsefly fire, unseen during training. Total pixel count is the total number of pixels tested. SSIM was the metric used for comparisons in this study.

Dataset	Total Pixel Count	SSIM
MASTER SIT-FUSE	20,389,114	0.8
MASTER Pre-Existing	20,389,114	0.8
GOES SIT-FUSE	1,016,785	0.82
GOES Pre-Existing	1,016,785	0.72
VIIRS SIT-FUSE	4,759,386	0.71
VIIRS Pre-Existing	4,759,386	0.59

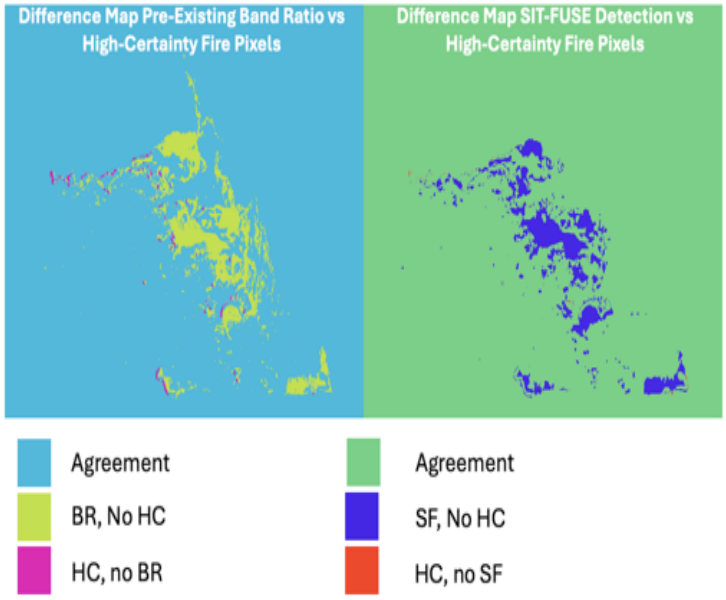


Figure 12. A pair of difference maps between different detection outputs and the hand-labeled high-certainty active fire pixels. The left panel is a difference map between the band-ratio-based detection (BR) and the high-certainty labels (HC). The right panel is a difference map between the SIT-FUSE-based detection (SF) and the high-certainty labels (HC).

3.2. Smoke Detection

Table 6 summarizes the performance of our SIT-FUSE-based segmentation approach, compared against the hand-labeled high-certainty smoke plume pixels, for plumes across all scenes in the test set, for each instrument tested. Table 7 presents the same specifically for smoke plumes in geographic areas not seen during training. In all cases, our approach performs well. When visually compared to the hand-labeled data, our approach tends to over-segment relative to the high-certainty smoke plume areas but performs well across the majority of the plume regions tested. The borders of the smoke plumes appear to provide uncertainty for both labelers and the SIT-FUSE-based automated detection. In some cases, these areas were not fully covered by our detections, whereas in other cases, the detections identified smoky areas in the scene that were not covered by the labels. Shadows also appear

to confuse small numbers of cases. Figure 13 provides an example detection produced for a MASTER scene capturing the Williams Flats fire, while Figure 14 shows the same example for an AirMSPI scene capturing the same active fire. Figure 15 shows difference maps over eMAS scenes of the same fire between the high-certainty hand labled data, SIT-FUSE smoke detections, and the eMAS AOD product, thresholded at 0.2 AOD. Our detection appears to capture more of the thick smoke plume, while not picking up areas surrouding the smoke plume that appear to be clear, but potentially have finer aerosols in the atmosphere.

Table 6. Summary of smoke detection comparisons against the hand-labeled high-certainty smoke plumes over the full set of test scenes that contain active fires. Total pixel count is the total number of pixels tested. SSIM was the metric used for comparisons in this study.

Dataset	Total Pixel Count	SSIM
MASTER	12,438,023	0.71
eMAS	162,644,645	0.63
AVIRIS-C	173,334,623	0.67
AirMSPI Sweep	132,238,386	0.63
PlanetScope Dove	1,353,594,326	0.68
GOES-17	1,016,785	0.62
VIIRS	4,759,386	0.66

Table 7. Summary of smoke detection comparisons against the hand-labeled high-certainty active fire areas over only the scenes that contain the Sheridan fire and Horsefly fire, unseen during training. Total pixel count is the total number of pixels tested. SSIM was the metric used for comparisons in this study. Planet data were not available over the Sheridan fire, so we cannot assess this.

Dataset	Total Pixel Count	SSIM
MASTER	1,708,746	0.75
eMAS	133,762,591	0.66
AVIRIS-C	21,073,834	0.77
AirMSPI Sweep	63,505,132	0.65
PlanetScope Dove	N/A	N/A
GOES-17	2,541,975	0.65
VIIRS	1,189,846	0.68

Table 8 summarizes the same set of comparisons, this time evaluating SIT-FUSE-based detections against results from operational dark-target-based detection methodologies. Both approaches are compared against the high-certainty active fire pixel label set for eMAS, GOES, and VIIRS. For these comparisons, we used the aerosol optical depth (AOD) value from the dark target product of each of the instruments and applied a minimum threshold value of 0.2.

Table 8. Summary of smoke detection comparisons against the hand-labeled high-certainty smoke plumes over only the scenes unseen during training. Total pixel count is the total number of pixels tested. SSIM was the metric used for comparisons in this study.

Dataset	Total Pixel Count	SSIM
eMAS SIT-FUSE	162,644,645	0.63
eMAS Pre-Existing	162,644,645	0.37
GOES SIT-FUSE	1,016,785	0.73
GOES Pre-Existing	1,016,785	0.56
VIIRS SIT-FUSE	4,759,386	0.66
VIIRS Pre-Existing	4,759,386	0.54

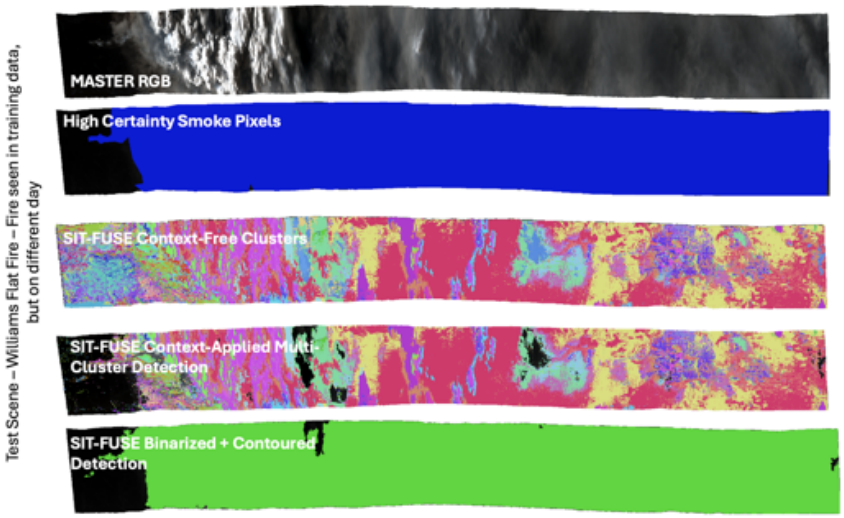


Figure 13. The top left panel is an RGB image extracted from a MASTER scene over the Williams Flats fire. Following this panel, on the top row are the hand-labeled high-certainty active fire pixels overlayed on the RGB image. The bottom row consists of the context-free segmentation map generated from SIT-FUSE for this scene; the context-applied multi-cluster detection, which is a subset of the full segmentation map with only the labels/clusters that correlate to smoke plumes; and the final binarized and contoured smoke mask.

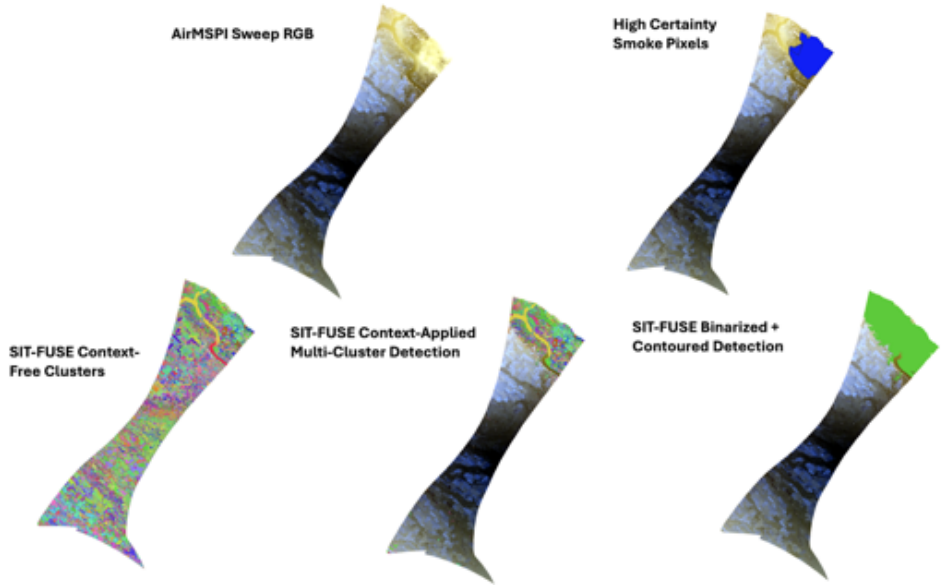


Figure 14. The top left panel is an RGB image extracted from an AirMSPI scene over the Williams Flats fire. Following this panel, on the top row are the hand-labeled high-certainty active fire pixels overlayed on the RGB image. The bottom row consists of the context-free segmentation map generated from SIT-FUSE for this scene; the context-applied multi-cluster detection, which is a subset of the full segmentation map with only the labels/clusters that correlate to smoke plumes; and the final binarized and contoured smoke mask.

3.3. Multi-Sensor Fusion

In previous work, we have demonstrated the utility of fusing multi-angle data from MISR with that of MODIS, both instruments onboard the same satellite platform, Terra. This included segmenting the Williams Flats fire and smoke plume in our evaluation. Figure 16 depicts segmentation, as well as active fire and smoke segmentation over the Williams Flats fire, using MISR and MODIS data as input. However, combining MISR and MODIS data is limited by the restricted number of overpasses for a single platform and the narrower area of swath intersection, which reduces the number of cases available for evaluation

compared to those in the single-instrument analyses described earlier. For polar-orbiting instruments, multiple platforms often provide an increased number of scenes for detection and evaluation. As we expand the spatiotemporal areas studied, we will add these fusion cases and quantify their benefits in a manner consistent with the previous evaluations.

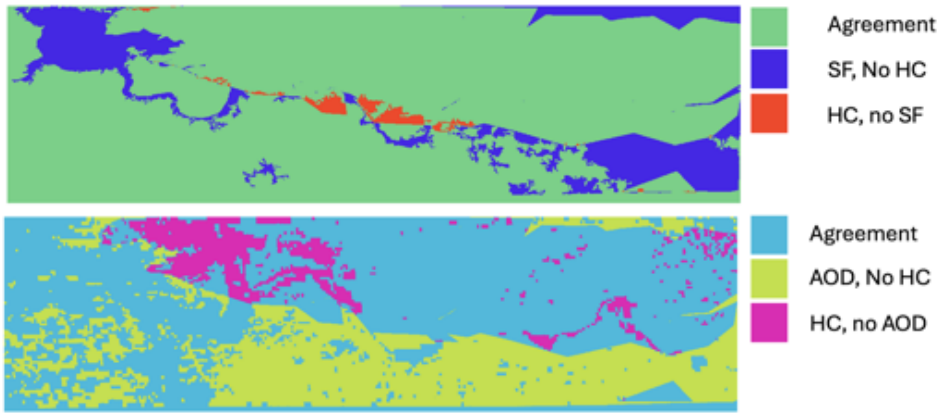


Figure 15. A pair of difference maps between different detection outputs and the hand-labeled high-certainty smoke pixels. The top panel is a difference map between the SIT-FUSE-based detection (SF) and the high-certainty labels (HC). The bottom panel is the difference between the operational eMAS AOD data, thresholded at 0.2 (AOD), and the high-certainty labels (HC).

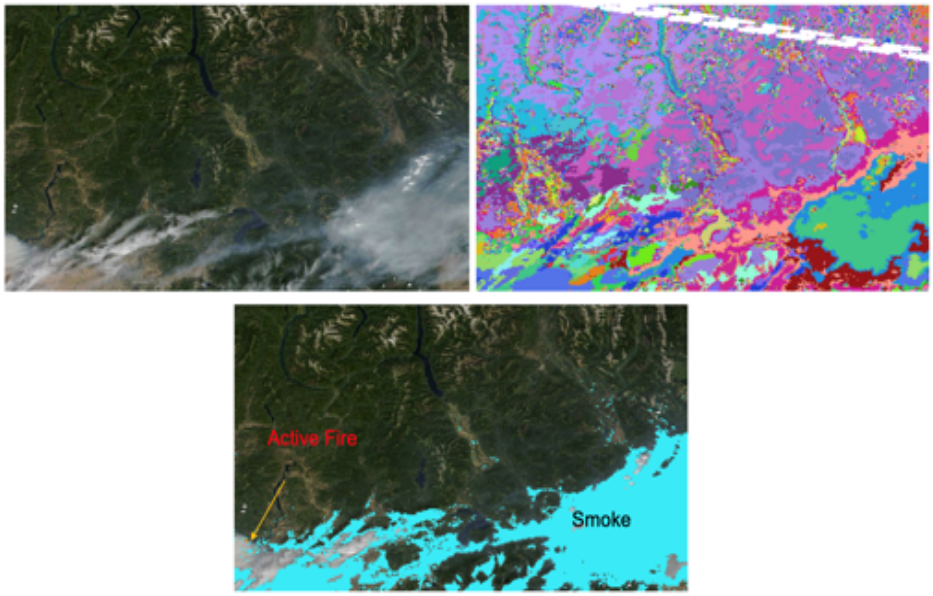


Figure 16. Example active wildfire front and smoke plume detection using Terra MODIS and 9-angle MISR scene over the Williams Flats fire on 8 August 2019, as input. The top left panel is a reference RGB generated from the MODIS data. The right panel is the context-free segmentation map generated from SIT-FUSE, and the bottom panel is the extracted and binarized smoke and active fire front detections from the SIT-FUSE outputs.

Given our previous success with the fusion of polar-orbiting instrument data, we explored fusing airborne instrumentation data, both within the same platform and across platforms. For the instruments on the same platform, we examined AirMSPI and eMAS, both onboard the ER-2 aircraft, as an airborne equivalent of MISR and MODIS. As a first step, we tested AirMSPI's multi-angle data taken in step-and-stare mode collectively as a simple fusion test case. Unfortunately, the coverage of the data is very small and the number of scenes is relatively low. While there is spatial overlap between data captured at

different viewing angles, data quality issues restrict their utility. There are areas marked as "bad/occluded data" that increased in frequency with more oblique angles. We tried to mitigate this by limiting the number of angles used, but it did not significantly improve, even after excluding data from four angles. Although we were able to generate smoke and active fire detection examples, the number of scenes and pixels was insufficient for both model training and separate evaluation. Figure 17 depicts an example scene from AirMSPI data, including the two most oblique angles and the nadir angle, as well as the minimal segmentation coverage we can obtain for SIT-FUSE when combining the multi-angle data, and the associated active fire and smoke masks.

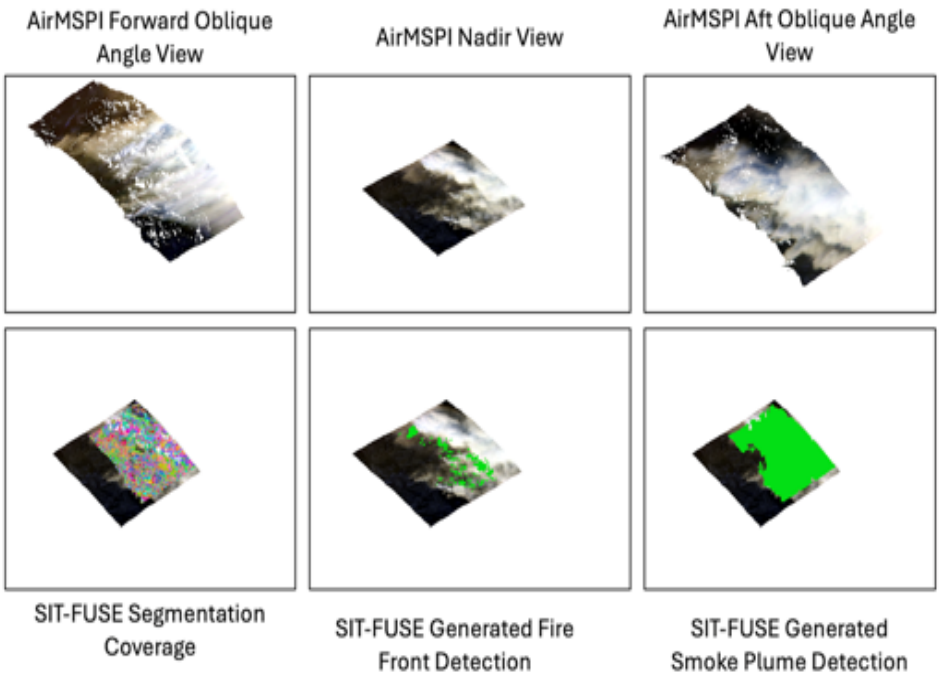


Figure 17. Example wildfire front and smoke plume detection using multi-angle AirMSPI step-and-stare data. The panels on the top row are RGBs from the two most oblique angles, split by the RGB generated from the nadir data from a scene over the Williams Flats fire on 6 August 2019. The bottom row depicts the full coverage for the SIT-FUSE segmentation map, overlayed on the nadir RGB for reference, the extracted active fire front detection, and the extracted smoke plume detection, respectively.

For the AirMSPI + eMAS case, we also attempted to combine the sweep mode data with eMAS data. However, the overlap between the datasets was minimal. Future campaigns may take into account the need to fuse data in this manner and attempt to collect a greater number of spatiotemporally collocated scenes with multiple instruments. Again, as we expand our case studies and datasets over larger areas and campaigns, we are confident that we can quantitatively assess the benefits of data fusion between AirMSPI and eMAS for active wildfire and smoke detection.

For the case of cross-platform fusion, we looked at fusing eMAS and MASTER. Again the number of scenes is limited, and only over the Williams Flats active fire area, but the initial examples demonstrate potential for this capability, and again future campaign planning and use of additional pre-existing campaign datasets will allow us to further validate these fusion cases. Figure 18 depicts an example of one of the fusion scenes.

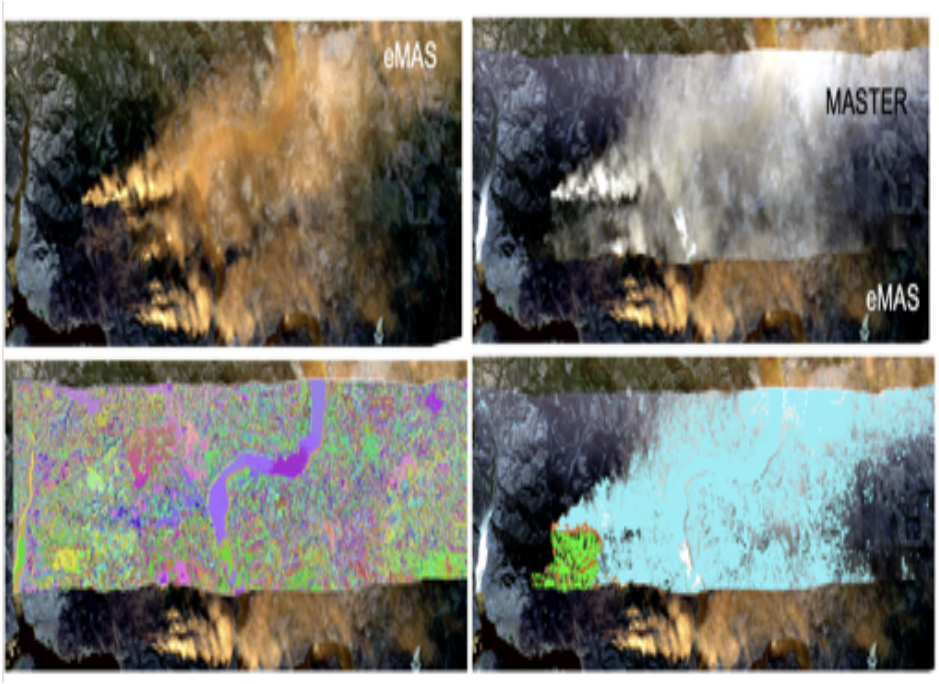


Figure 18. Example active wildfire front and smoke plume detection using MASTER and eMAS data as input. The top row depicts an RGB from an eMAS scene on 6 August 2019, and an RGB generated from the collocated MASTER scene overlayed on the eMAS RGB, both of which are used as input. The bottom row contains the context-free segmentation map overlayed on the area that contains both MASTER and eMAS data, and the extracted and binarized smoke and active fire front detections from the SIT-FUSE outputs.

3.4. Commercial Remote Sensing Products

As a part of this effort, we evaluated the utility of PlanetScope data for active wildfire and smoke detection and monitoring. We found that the older PlanetScope SuperDove imagers available during the FIREX-AQ 2019 campaign were not suited for identifying active fire fronts across the entire test set. However, they were effective at detecting smoke within the scenes evaluated. Figure 19 depicts smoke detection within a mosaic of SuperDove scenes over the Williams Flats fire on 8 August 2019, and Table 6 contains quantitative evaluation results. The improved temporal resolution of smoke detections at high spatial resolution will be helpful both for air quality research and automated smoke tracking. Also, as Planet's instrument offerings continue to diversify, with new SuperDove instruments having a higher spectral resolution, hyperspectral instrumentation being launched, and other companies continuing to launch various other Earth-orbiting instrumentation, it is critical that we look at the tradeoffs between paying for data from commercial entities and the scientific value added by such data. In this study, we analyzed 83 additional scenes over the Williams Flats area over four days. Per the NASA/Planet CSDSA agreement, we are not able to release the input SuperDove data, alongside all of the other data we are releasing. However, we will release the generated output and all other associated data.

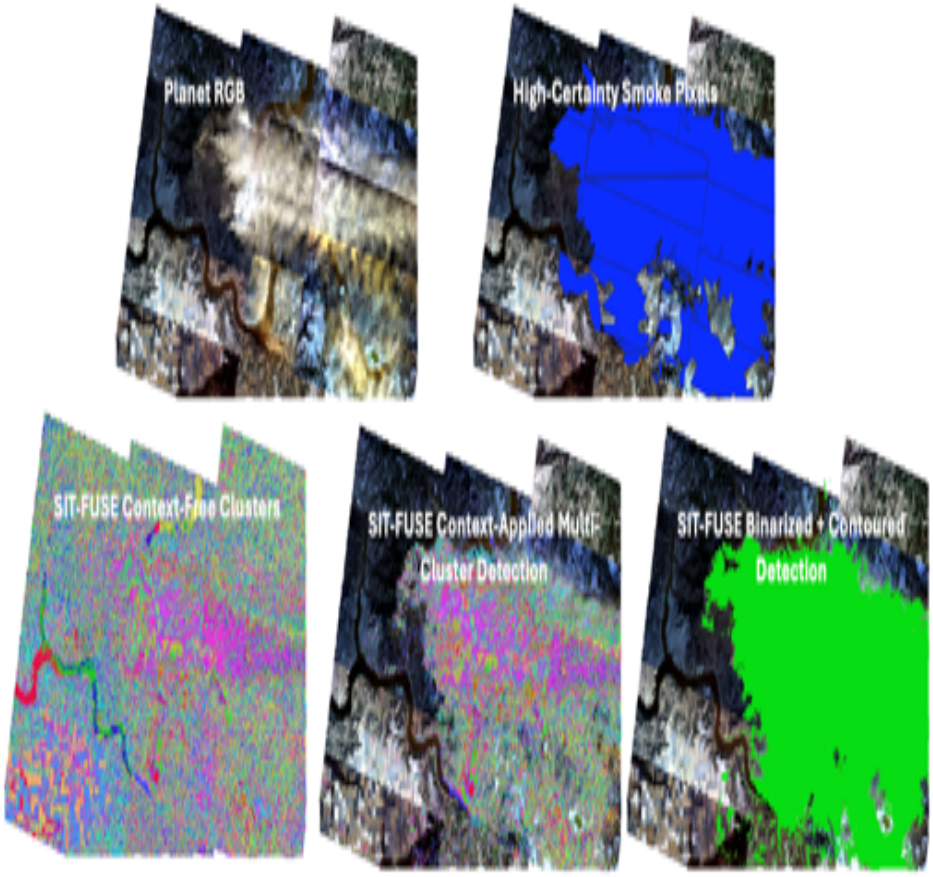


Figure 19. An example of using Planet data to perform smoke detection. The top row depicts a mosaic of RGBs generated from the input scenes and the high-certainty smoke pixels overlayed on top. The bottom row contains the context-free segmentation map generated from SIT-FUSE, the subset labels/clusters that are associated with smoke, and the final panel contains the contoured and binarized smoke plume detected over the entire mosaic. Image © 2019 Planet Labs PBC.

4. Conclusions

Overall, SIT-FUSE effectively identifies and segments wildland active fire fronts and smoke plumes, and performs better against high-certainty smoke and active fire label sets, when compared against other operational and experimental approaches, as seen in Tables 5 and 8. As expected, the higher the resolution of the instrument, the more detailed segmentation can occur. However, the different resolutions are valuable not only as standalone datasets but also for tracking these objects across multiple datasets. While further large-scale validation is needed, our tests demonstrate the capability to increase the temporal resolution of active fire front products. We achieve this by increasing the number of instruments available for active fire and smoke segmentation and developing brand-new active fire front and smoke plume products for many of the instruments tested. This indicates significant potential for both product generation and utilization of this technique for segmentation and instance tracking. This not only also allows for dynamic instance tracking across scenes with the same input set, but we believe by harnessing style transfer capabilities, we can also look into instance tracking across multi-sensor scenes from disparate input datasets. We hope to transition this technology into a piece of future operational campaign support or product generation systems.

5. Discussion and Current/Future Work

In terms of feature interpretability and selection, methods such as SHAP analysis and other explainability methods can be applied to better understand feature importance and refine the input to focus on spectral bands most effective for identifying smoke and/or active fire. Given the current performance and the success with datasets where there was no pre-existing operational active fire or smoke detection methodology, solutions like SIT-FUSE can be integrated into new or existing instrumentation data processing pipelines. By doing so, this approach could replace or augment instrument-specific retrieval algorithms, which may be extremely costly to develop. SIT-FUSE's segmentation capabilities offer additional benefits: the decrease in data volume processed for downstream active fire- or smoke-specific retrievals. By isolating the detected objects, only relevant pixels need to be processed through a downstream retrieval, thereby optimizing the pipeline.

We have built a framework within SIT-FUSE that is adaptable to various kinds of encoders and we aim to be able to leverage this to analyze representative capabilities of different model types, complexities, and training paradigms. With the continued influx of new architectures and large Earth Observation Foundation Models, it is important to understand these models provide quality representations (or poor ones) under different conditions, problem sets, and input datasets [68]. Analyses of downstream task performance are a crucial piece, but not the entire solution. More robust ways to evaluate representative capabilities are emerging around large language models (LLMs), and much of this can be ported to computer vision, and specifically deep learning for Earth observations [69]. Within the flexible framework of SIT-FUSE, we are working towards providing initial pathways towards tackling some of these open problems.

Lastly, we are working to leverage SIT-FUSE to make an impact within the area of analysis and scientific understanding—in this case, correlated to active wildfires and smoke plumes. There is a built-in co-discovery facilitation mechanism, by way of the hierarchical context-free segmentation products. By using the model-derived separations of various areas, novelty and "interesting" samples can more easily be grouped and investigated. This can be even further coupled with more detailed analyses of the embedding spaces relative to the context-free segmentations [3]. To enhance exploration even further, models trained for co-exploration of data using open-ended algorithms can be leveraged to more quickly sift through the volumes of data and highlight interesting, new, and anomalous samples [70,71].

Author Contributions: Conceptualization, N.L., K.Y., M.J.G. and E.L.; Data curation, N.L., A.E., H.L., M.J.G. and O.V.K.; Formal analysis, N.L. and A.E.; Funding acquisition, N.L., K.Y., H.L., M.J.G. and O.V.K.; Investigation, N.L., K.Y., M.J.G., O.V.K. and E.L.; Methodology, N.L., K.Y., H.L. and E.L.; Project administration, N.L.; Software, N.L., K.Y., H.L. and E.L.; Supervision, N.L., M.J.G., O.V.K. and E.L.; Validation, N.L., K.Y., H.L., M.J.G., O.V.K. and E.L.; Visualization, N.L., K.Y., H.L. and E.L.; Writing—original draft, N.L.; Writing—review and editing, N.L., K.Y., H.L., M.J.G., O.V.K. and E.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). This research was funded by the NASA ROSES Commercial Smallsat Data Scientific Analysis program (NNH22ZDA001N-CSDSA) as well as the Jet Propulsion Laboratory's Data Science Working Group. Computing resources were leveraged at both the NASA Center for Climate Simulation (NCCS) and the Machine Learning and Affiliated Technologies (MLAT) Lab in the Fowler School of Engineering at Chapman University.

Data Availability Statement: The data have been published and are freely available on Zenodo [72]. The model weights and associated configuration files are also publicly available on HuggingFace [73]. Lastly, the code has been tagged at the time of this paper submission and is publicly available

on GitHub [50]. The Planet radiance data cannot be released publicly due to the terms of the NASA Commercial Smallsat Data Acquisition agreement, but all products generated and models trained have been shared, and all other input data are freely available in the aforementioned Zenodo data store.

Acknowledgments: The authors would like to thank NASA, JPL, NCCS, MLAT Lab., The Spatial Informatics Group, LLC., and the Schmid College of Science and Technology, Chapman University, for supporting this research. The authors would also send thanks to Phil Dennison, from the Geography Department, at the University of Utah, as well as NASA for providing the data, without which this research would not have been possible. Finally, the authors would like to thank the anonymous reviewers for taking the time to read this paper and provide valuable feedback.

Conflicts of Interest: Authors Nicholas LaHaye was employed by the company Spatial Informatics Group LLC. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. LaHaye, N.; Ott, J.; Garay, M.J.; El-Askary, H.M.; Linstead, E. Multi-Modal Object Tracking and Image Fusion With Unsupervised Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3056–3066. [CrossRef]
2. LaHaye, N.; Garay, M.J.; Bue, B.D.; El-Askary, H.; Linstead, E. A Quantitative Validation of Multi-Modal Image Fusion and Segmentation for Object Detection and Tracking. *Remote Sens.* **2021**, *13*, 2364. [CrossRef]
3. LaHaye, N.; Lee, H.; Easley, A.; Garay, M.J.; Yun, K.; Goodman, A.; Kalashnikova, O.V. A Comparison of Model Complexity, Representative Capabilities, and Performance for Self-Supervised Multi-Sensor Wildifre and SMoke Segmentation and Tracking. In Proceedings of the AGU Fall Meeting 2024. American Geophysical Union, Washington, DC, USA, 9–13 December 2024.
4. Warneke, C.; Schwarz, J.P.; Dibb, J.; Kalashnikova, O.; Frost, G.; Al-Saad, J.; Brown, S.S.; Brewer, W.A.; Soja, A.; Seidel, F.C.; et al. Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ). *J. Geophys. Res. Atmos.* **2023**, *128*, e2022JD037758. [CrossRef]
5. Williams, A.P.; Abatzoglou, J.T.; Gershunov, A.; Guzman-Morales, J.; Bishop, D.A.; Balch, J.K.; Lettenmaier, D.P. Observed Impacts of Anthropogenic Climate Change on Wildfire in California. *Earth Future* **2019**, *7*, 892–910. [CrossRef]
6. Stavros, N.; Agha, A.; Alkalai, L.; Sirota, A.; Quadrelli, M.; Ebadi, K.; Yun, K. *Smoke Sky: Exploring New Frontiers of Unmanned Aerial Systems for Wildland Fire Science and Applications*; CSIRO: Pasadena, CA, USA, 2019. [CrossRef]
7. Barmpoutis, P.; Papaioannou, P.; Dimitropoulos, K.; Grammalidis, N. A Review on Early Forest Fire Detection Systems Using Optical Remote Sensing. *Sensors* **2020**, *20*, 6442. [CrossRef] [PubMed]
8. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [CrossRef]
9. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988. [CrossRef]
10. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *arXiv Prepr.* **2021**, arXiv:2105.15203.
11. King, M.D.; Kaufman, Y.J.; Menzel, W.P.; Tanre, D. Remote Sensing of Cloud, Aerosol, and Water Vapor Properties from the Moderate Resolution Imaging Spectrometer (MODIS). *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 2–27. [CrossRef]
12. GIBS, N. Worldview Snapshots—wvs.earthdata.nasa.gov. Available online: https://wvs.earthdata.nasa.gov (accessed on 24 December 2024).
13. Knight, E.; Kvaran, G. Landsat-8 Operational Land Imager Design, Characterization and Performance. *Remote Sens.* **2014**, *6*, 10286–10305. [CrossRef]
14. Kanezaki, A. Unsupervised Image Segmentation by Backpropagation. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 1543–1547. [CrossRef]
15. Chen, M.; Artières, T.; Denoyer, L. Unsupervised Object Segmentation by Redrawing. *arXiv* **2019**, arXiv:1905.13539.
16. Aganj, I.; Harisinghani, M.G.; Weissleder, R.; Fischl, B. Unsupervised Medical Image Segmentation Based on the Local Center of Mass. *Sci. Rep.* **2018**, *8*, 13012. [CrossRef]
17. Soares, A.R.; Körting, T.S.; Fonseca, L.M.G.; Neves, A.K. An Unsupervised Segmentation Method For Remote Sensing Imagery Based on Conditional Random Fields. In Proceedings of the 2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS), Santiago, Chile, 22–26 March 2020; pp. 1–5. [CrossRef]

18. Zhang, R.; Yu, L.; Tian, S.; Lv, Y. Unsupervised remote sensing image segmentation based on a dual autoencoder. *J. Appl. Remote Sens.* **2019**, *13*, 038501. [CrossRef]
19. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything. *arXiv* **2023**, arXiv:2304.02643.
20. Thangavel, K.; Spiller, D.; Sabatini, R.; Amici, S.; Sasidharan, S.T.; Fayek, H.; Marzocca, P. Autonomous Satellite Wildfire Detection Using Hyperspectral Imagery and Neural Networks: A Case Study on Australian Wildfire. *Remote Sens.* **2023**, *15*, 720. [CrossRef]
21. Spiller, D.; Carbone, A.; Amici, S.; Thangavel, K.; Sabatini, R.; Laneve, G. Wildfire Detection Using Convolutional Neural Networks and PRISMA Hyperspectral Imagery: A Spatial-Spectral Analysis. *Remote Sens.* **2023**, *15*, 4855. [CrossRef]
22. Boroujeni, S.P.H.; Razi, A.; Khoshdel, S.; Afghah, F.; Coen, J.L.; O'Neill, L.; Fule, P.; Watts, A.; Kokolakis, N.M.T.; Vamvoudakis, K.G. A comprehensive survey of research towards AI-enabled unmanned aerial systems in pre-, active-, and post-wildfire management. *Inf. Fusion* **2024**, *108*, 102369. [CrossRef]
23. Real-Time Tracking of Wildfire Boundaries Using Satellite Imagery—Research.Google. Available online: https://research.google/blog/real-time-tracking-of-wildfire-boundaries-using-satellite-imagery/ (accessed on 24 December 2024).
24. Sun, Y.; Jiang, L.; Pan, J.; Sheng, S.; Hao, L. A satellite imagery smoke detection framework based on the Mahalanobis distance for early fire identification and positioning. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *118*, 103257. [CrossRef]
25. Diner, D.J.; Xu, F.; Garay, M.J.; Martonchik, J.V.; Rheingans, B.E.; Geier, S.; Davis, A.; Hancock, B.R.; Jovanovic, V.M.; Bull, M.A.; et al. The Airborne Multiangle SpectroPolarimetric Imager (AirMSPI): A new tool for aerosol and cloud remote sensing. *Atmos. Meas. Tech.* **2013**, *6*, 2007–2025. [CrossRef]
26. Guerin, D.C.; Fisher, J.; Graham, E.R. The enhanced MODIS airborne simulator hyperspectral imager. In Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVII, Orlando, FL, USA, 20 May 2011; Volume 8048, pp. 214–224. [CrossRef]
27. Hook, S.J.; Myers, J.J.; Thome, K.J.; Fitzgerald, M.; Kahle, A.B. The MODIS/ASTER airborne simulator (MASTER)—a new instrument for earth science studies. *Remote Sens. Environ.* **2001**, *76*, 93–102. [CrossRef]
28. Green, R.O.; Eastwood, M.L.; Sarture, C.M.; Chrien, T.G.; Aronsson, M.; Chippendale, B.J.; Faust, J.A.; Pavri, B.E.; Chovit, C.J.; Solis, M.; et al. Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sens. Environ.* **1998**, *65*, 227–248. [CrossRef]
29. Diner, D. *MISR Experiment Overview*; NASA: Washington, DC, USA, 1999.
30. U.S. Department of Commerce. *Visible Infrared Imaging Radiometer Suite (VIIRS) Sensor Data Record (SDR) User's Guide*; U.S. Department of Commerce: Washington, DC, USA, 2017.
31. DOC; NOAA; NESDIS; NASA. *GOES-R Series Product Definition and Users' Guide*; NASA: Washington, DC, USA, 2019.
32. Planet, T. *Planet Application Program Interface: In Space for Life on Earth*; Planet Development Center: San Francisco, CA, USA, 2018.
33. MacQueen, J.B. Some Methods for Classification and Analysis of MultiVariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*; Cam, L.M.L., Neyman, J., Eds.; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–297.
34. Lin, B.; Yu, Z.; Huang, F.; Guo, L. Pool-Based Sequential Active Learning For Regression Based on Incremental Cluster Center Selection. In Proceedings of the 2021 Ninth International Conference on Advanced Cloud and Big Data (CBD), Xi'an, China, 26–27 March 2021; pp. 176–182.
35. Burke, A.; Carroll, M.; Spradlin, C. Finding the Trees in a (Random) Forest: How Do We Get a Representative Sample in a Training Dataset for a Global Land Cover Classification? In Proceedings of the AGU Fall Meeting Abstracts, San Francisco, CA, USA, 11–15 December 2023; Volume 2023, p. IN51C–0429.
36. Grill, J.B.; Strub, F.; Altché, F.; Tallec, C.; Richemond, P.H.; Buchatskaya, E.; Doersch, C.; Pires, B.A.; Guo, Z.D.; Azar, M.G.; et al. Bootstrap your own latent: A new approach to self-supervised Learning. *arXiv* **2020**, arXiv:2006.07733.
37. Assran, M.; Duval, Q.; Misra, I.; Bojanowski, P.; Vincent, P.; Rabbat, M.G.; LeCun, Y.; Ballas, N. Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. *arXiv* **2023**, arXiv:2301.08243.
38. Liao, R.; Kornblith, S.; Ren, M.; Fleet, D.J.; Hinton, G. Gaussian-Bernoulli RBMs Without Tears. *arXiv* **2022**, arXiv:2210.10318.
39. He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; Girshick, R.B. Masked Autoencoders Are Scalable Vision Learners. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 15979–15988.
40. de Boves Harrington, P. Feature expansion by a continuous restricted Boltzmann machine for near-infrared spectrometric calibration. *Anal. Chim. Acta* **2018**, *1010*, 20–28. [CrossRef]
41. Sobczak, S.; Kapela, R. Restricted Boltzmann Machine as Image Pre-processing Method for Deep Neural Classifier. In Proceedings of the 2019 First International Conference on Societal Automation (SA), Krakow, Poland, 4–6 September 2019; pp. 1–5. [CrossRef]
42. Harrington, P.B. Enhanced zippy restricted Boltzmann machine for feature expansion and improved classification of analytical data. *J. Chemom.* **2020**, *34*, e3228. [CrossRef]

43. Ji, X.; Vedaldi, A.; Henriques, J.F. Invariant Information Clustering for Unsupervised Image Classification and Segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9864–9873.
44. Suzuki, S.; Be, K. Topological structural analysis of digitized binary images by border following. *Comput. Vision Graph. Image Process.* **1985**, *30*, 32–46. [CrossRef]
45. Bradski, G. The OpenCV Library. In *Dr. Dobb's Journal of Software Tools*; The University of Utah: Salt Lake, UT, USA, 2000.
46. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *Image Process. IEEE Trans.* **2004**, *13*, 600–612. [CrossRef]
47. Wang, Z.; Bovik, A.C. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Process. Mag.* **2009**, *26*, 98–117. [CrossRef]
48. Woodhouse, I.H. On 'ground' truth and why we should abandon the term. *J. Appl. Remote Sens.* **2021**, *15*, 041501. [CrossRef]
49. Yang, F.; Du, M.; Hu, X. Evaluating Explanation Without Ground Truth in Interpretable Machine Learning. *arXiv* **2019**, arXiv:1907.06831.
50. LaHaye, N.; Easley, A.; Leet, N. *SIT_FUSE*. 2025. [CrossRef]
51. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef]
52. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
53. Dask Development Team. Dask; Dask Development Team. 2016. Available online: https://dask-local.readthedocs.io/en/latest/cite.html (accessed on 10 August 2024).
54. Hoyer, S.; Hamman, J. xarray: N-D labeled arrays and datasets in Python. *J. Open Res. Softw.* **2017**, *5*, 148. [CrossRef]
55. Zenodo. *zarr-developers/zarr-python, Version 3.0.1*; Zenodo/CERN: Geneva, Switzerland, 2025. [CrossRef]
56. Lam, S.K.; Pitrou, A.; Seibert, S. Numba: A LLVM-based Python JIT compiler. In Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, New York, NY, USA, 15 November 2015; LLVM '15. [CrossRef]
57. Okuta, R.; Unno, Y.; Nishino, D.; Hido, S.; Loomis, C. CuPy: A NumPy-Compatible Library for NVIDIA GPU Calculations. In Proceedings of the Workshop on Machine Learning Systems (LearningSys) in the Thirty-first Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
58. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 8024–8035.
59. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
60. Roder, M.; de Rosa, G.H.; Papa, J.P. Learnergy: Energy-based Machine Learners. *arXiv* **2020**, arXiv:2003.07443.
61. Zenodo. *pytroll/pyresample, Version 1.31.0*; Zenodo/CERN: Geneva, Switzerland, 2024. [CrossRef]
62. Rouault, E.; Warmerdam, F.; Schwehr, K.; Kiselev, A.; Butler, H.; Łoskot, M.; Szekeres, T.; Tourigny, E.; Landa, M.; Miara, I.; et al. *GDAL*; Zenodo/CERN: Geneva, Switzerland, 2025. [CrossRef]
63. Zonca, A.; Singer, L.; Lenz, D.; Reinecke, M.; Rosset, C.; Hivon, E.; Gorski, K. healpy: Equal area pixelization and spherical harmonics transforms for data on the sphere in Python. *J. Open Source Softw.* **2019**, *4*, 1298. [CrossRef]
64. *ssec/polar2grid, Python Package Version 3.2.0*; Zenodo/CERN: Geneva, Switzerland, 2025. [CrossRef]
65. Jordahl, K.; den Bossche, J.V.; Fleischmann, M.; Wasserman, J.; McBride, J.; Gerard, J.; Tratner, J.; Perry, M.; Badaracco, A.G.; Farmer, C.; et al. *geopandas/geopandas, Version 0.8.1*; Zenodo/CERN: Geneva, Switzerland, 2020. [CrossRef]
66. Itseez. Open Source Computer Vision Library. 2015. Available online: https://github.com/itseez/opencv (accessed on 15 January 2025).
67. QGIS Development Team. *QGIS Geographic Information System*; Open Source Geospatial Foundation Project: Beaverton, OR, USA, 2024.
68. Marsocci, V.; Jia, Y.; Bellier, G.L.; Kerekes, D.; Zeng, L.; Hafner, S.; Gerard, S.; Brune, E.; Yadav, R.; Shibli, A.; et al. PANGAEA: A Global and Inclusive Benchmark for Geospatial Foundation Models. *arXiv* **2024**, arXiv:2412.04204.
69. Duderstadt, B.; Helm, H.S.; Priebe, C.E. Comparing Foundation Models using Data Kernels. *arXiv* **2023**, arXiv:2305.05126.
70. Zhang, J.; Lehman, J.; Stanley, K.; Clune, J. OMNI: Open-endedness via Models of human Notions of Interestingness. *arXiv* **2023**, arXiv:2306.01711.
71. Lu, C.; Lu, C.; Lange, R.T.; Foerster, J.; Clune, J.; Ha, D. The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery. *arXiv* **2024**, arXiv:2408.06292.

72. LaHaye, N.; Easley, A.; Lee, H.; Yun, K.; Linstead, E.; Garay, M.; Kalashnikova, O. *Multi-Platform/Multi-Sensor Fire and Smoke Segmentation from the FIREX-AQ 2019 Campaign*; Zenodo/CERN: Geneva, Switzerland, 2025. [CrossRef]
73. LaHaye, N. *FIREX_AQ_2019_Fire_Smoke (Revision e2f8a55)*; HuggingFace: New York, NY, USA, 2025. [CrossRef]