TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Traba, C. C., Rijlaarsdam, D., Guo, J., Del Prete, R., & Meoni, G. (2026). Towards onboard thermal hotspots segmentation with raw multispectral satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, *146*, Article 105095. https://doi.org/10.1016/j.jag.2026.105095

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

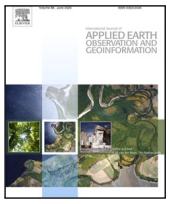
This work is downloaded from Delft University of Technology.

International Journal of Applied Earth Observation and Geoinformation 146 (2026) 105095

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Towards onboard thermal hotspots segmentation with raw multispectral satellite imagery

Cristopher Castro Traba [a,b,c], David Rijlaarsdam [b], Jian Guo [a,*], Roberto Del Prete [c], Gabriele Meoni [c,d]

[a] Delft University of Technology, Kluyverweg 1, Delft, 2629 HS, South Holland, Netherlands
[b] Ubotica Technologies, Kluyverweg 1, Delft, 2629 HS, South Holland, Netherlands
[c] Φ-lab, European Space Research Institute (ESRIN), European Space Agency (ESA), Via Galileo Galilei, 1, Frascati, 00044, RM, Italy
[d] Advanced Concepts and Studies Office, European Space Research Institute (ESRIN), European Space Agency (ESA), Via Galileo Galilei, 1, Frascati, 00044, RM, Italy

ARTICLE INFO

Dataset link: https://github.com/Ubotica/SegTHRawS_ext, https://zenodo.org/records/14741990

Keywords:
Onboard AI
Wildfire
Volcanic eruption
Online processing
AI4EO
EO satellite
Edge computing

ABSTRACT

The rapid spread and destructive nature of wildfires and volcanic activity have intensified the need for low latency detection systems. The growing intensity and frequency of globally distributed thermal hotspots have driven the development of satellite-based detection solutions. Conventional approaches rely on ground-based processing, which limits low latency capabilities due to revisit times over ground stations and data handling requirements. This work proposes the first onboard payload processing pipeline for segmentation of thermal hotspots in raw multispectral satellite imagery. The pipeline leverages the Near InfraRed (NIR) and Short-Wave InfraRed (SWIR) spectral bands, and the combination of onboard Artificial Intelligence (AI) and raw imagery significantly reduces the delay between image acquisition and event detection. Furthermore, we present Segmentation of Thermal Hotspots in Raw Sentinel-2 data (SegTHRawS), the first publicly available dataset for thermal hotspot segmentation in raw multispectral satellite imagery. The segmentation model employed is a Fully Convolutional Network (FCN) derived from U-Net, named ResUnet-S2, designed for fast on-device inference. This model achieved an Intersection over Union (IoU) of 0.988 and an F-1 score of 0.986 on SegTHRawS, with its detection and generalization capabilities validated using an external thermal hotspot segmentation dataset. The proposed pipeline was verified on CubeSat-compatible hardware, achieving an end-to-end execution, from image acquisition to event detection, in 1.45 s, faster than the image acquisition process, and consuming a peak power of 4.05 W. These results demonstrate the potential of onboard processing solutions for minimizing the detection latency of current approaches, particularly for thermal hotspot segmentation, using edge computing satellite hardware.

1. Introduction

Uncontrolled thermal hostpots or anomalies, such as wildfires and volcanic activity, have substantial social, environmental, and economic impacts on the environment and human welfare (Bowman et al., 2017; Wooster et al., 2021). These events generate long-term consequences that impact wildlife and vegetation, air pollution, weather patterns, and global warming (Toan et al., 2019; Seydi et al., 2022). Longer and more severe fire seasons and extreme volcanic activity in the last decades show the importance of thermal hotspot detection and monitoring (Tyukavina et al., 2022). For example, the 2018 Camp Fire in California, known as the deadliest wildfire in California, burned over 60,000 hectares (600 km^2) causing 85 civilian fatalities, and approximately $16.5 billion in property damage (Schulze et al., 2020). Due to the harmful nature of wildfires and volcanic eruptions, these events need to be precisely detected as soon as they occur before being beyond control (Toan et al., 2019). The global distribution of thermal hotspots has driven a shift towards Earth Observation (EO) satellite-based monitoring solutions. These systems provide the necessary spatial coverage for global monitoring and their extended mission lifetimes enable operational capabilities for several years. Nevertheless, the temporal resolution of current satellite systems, characterized by limited revisit times, poses a significant challenge to the timely detection of thermal events, thereby affecting the latency between image acquisition, event detection, and communication to the emergency services.

Existing large-scale solutions, such as Fire Information for Resource Management System (FIRMS) (Davies et al., 2008) for wildfires and

* Correspondence to: Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands.
E-mail address: j.guo@tudelft.nl (J. Guo).

https://doi.org/10.1016/j.jag.2026.105095
Received 10 April 2025; Received in revised form 5 January 2026; Accepted 5 January 2026
Available online 15 January 2026
1569-8432/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Middle InfraRed Observation of Volcanic Activity (MIROVA) for volcanic activity (Coppola et al., 2016), try to provide Near Real-Time (NRT) monitoring through a multi-sensor approach that combines imagery from multiple satellite missions. These systems offer varying latencies: MIROVA provides NRT monitoring with a 1-3 h delay globally, whereas FIRMS provides Ultra Real-Time (URT) 60-second monitoring to limited areas in North America, NRT monitoring in North America, and ~3-hour delay in the rest of the world. These solutions use contextual and threshold detection methods that rely on the Thermal InfraRed (TIR) spectral bands with spatial resolutions ranging from hundreds to thousands of meters. These methods are based on manually crafted features, typically defined by trial and error (Wright et al., 2002; Pu et al., 2004). Their performance is highly dependent on factors such as background geography, time of day, season, and varying environmental conditions within the observed area (Wright et al., 2002), as radiance is influenced by the observed material (Schroeder et al., 2014; Giglio et al., 2016). This dependency limits the applicability of threshold-based methods, as specific thresholds must be defined for each unique scenario (Seydi et al., 2022). As an example to overcome this challenge, Hu et al. (2021a) introduced a multi-criteria approach known as AFD-S2, which employs statistical regression to determine dynamic thresholds across different biomes for active fire detection. However, this method lacks scalability, considers only a limited subset of biomes, and is restricted to wildfires. AI-based methods present a promising alternative to overcome these challenges (Seydi et al., 2022; Del Rosso et al., 2021; de Almeida Pereira et al., 2021; Meoni et al., 2025). These approaches enable the automatic extraction of image features, facilitating the identification of thermal events without requiring predefined thresholds.

The passes over ground stations limit the real-time performance of these solutions. The satellite is not always visible by a ground station, which leads to longer times between image acquisition and event detection. Additionally, current approaches require all information to be communicated at each ground station pass, potentially requiring more than one ground station pass, and then decompressing and processing the data to detect the event. On-board approaches can reduce the existing delay between image acquisition and even detection and reduce the amount of data to be transmitted to the ground (Yu et al., 2020; Giuffrida et al., 2020; Melega et al., 2023; Zhang et al., 2022). These concepts can be seen in the CogniSAT-6 mission (Rijlaarsdam et al., 2025), which proposed an innovative onboard solution capable of detecting an event, communicating insights through an Inter-Satellite-Link (ISL), and receiving them on Earth within minutes from image acquisition. On-board event detection optimizes data transmission as only extracted insights need to be communicated, reducing the required time to download, decompress, and process the data.

To further reduce detection latency, approaches leverage raw data or data with minimal pre-processing. This study adopts the definition of raw EO data from Meoni et al. (2024), encompassing decompressed (Level-0) imagery and associated metadata, excluding radiometric and geometric corrections. Removing computationally intensive image pre-processing steps substantially reduces detection delays and enables their implementation on resource-constrained space platforms, such as CubeSats (Bretschneider et al., 2005). Prior to the Thermal Hotspots in Raw Sentinel-2 data (THRawS) dataset (Meoni et al., 2024), research on thermal hotspot detection in raw multispectral imagery was absent, likely attributed to the lack of available datasets.

The complex nature of raw EO imagery poses substantial challenges for EO applications development. Without radiometric calibration, images are represented in raw Digital Numbers (DN), which include sensor-induced radiometric noise, rather than in reflectance values necessary for accurate object identification. Multispectral push-broom sensors measure a single band at a time per detector over a specific area, leading to temporal misregistration among bands due to the satellite's orbital movement (Chen and Liu, 2021). This issue is particularly evident in the Sentinel-2 mission, where the specific detector arrangement caused additional along-track band displacements up to 17 km (Gascon et al., 2017), resulting in misalignments in the order of hundreds of pixels within the image (Meoni et al., 2024).

To address these limitations, this study proposes the concept of an onboard AI-based payload processing pipeline for thermal hotspot segmentation using raw multispectral satellite imagery. By adopting an onboard processing design, the pipeline extracts actionable insights directly in-orbit, thereby minimizing the total time between image acquisition and thermal hotspot detection. Developing such a system, however, is hindered by the lack of public datasets for training the Deep Learning (DL) segmentation model. To overcome this challenge, another contribution of this work is the creation of the first publicly available segmentation dataset for thermal hotspots in raw Sentinel-2 imagery.

The main contributions of this paper are as follows:

- Design the concept of a processing pipeline with on-orbit capabilities for segmentation of thermal hotspots in satellites with limited power and processing capabilities. It was verified with edge computing CubeSat-compatible hardware, being comprehensively tested for processing time, power, and detection performance.
- Creation, to the best of the authors' knowledge, of the first publicly available dataset for thermal hotspot segmentation in raw multispectral satellite imagery, particularly for the Sentinel-2 mission.
- Creation of a contextual algorithm for thermal hotspots segmentation in multispectral imagery expressed in raw DNs.

The remainder of this article presents the following structure. Section 2 describes the related work in current solutions for thermal hotspot detection, and existing onboard approaches and concepts. Section 3 presents the methodology which describes the proposed onboard pipeline, the segmentation dataset creation process, the fast-inference segmentation models proposed for onboard implementation, the training of the DL models, the validation of the designed networks in an external dataset, and the experiment to verify the payload processing pipeline in embedded CubeSat hardware. Section 4 presents the results of the generated dataset, the trained models, and the experimental results of the pipeline in the selected hardware. Section 5 analyses and discusses the results, and Section 6 concludes the study.

2. Related work

Satellite imagery has been actively researched for the detection of wildfires and volcanic activity over the past several decades. Despite its extensive research, there remains a notable gap in the development of solutions specifically designed for low latency detection of such events. Existing approaches predominantly emphasize detection methodologies, often neglecting critical factors such as processing time requirements and the practical constraints associated with implementing these solutions in satellite missions. Low latency detection solutions can be broadly categorized into two approaches: (1) large-scale systems that integrate imagery from multiple missions to detect thermal events in real time, and (2) end-to-end small-scale solutions specifically designed for a particular satellite mission.

Large-scale solutions include systems like FIRMS (Davies et al., 2008) and MIROVA (Coppola et al., 2016) that combine images from several satellite missions to provide NRT global monitoring of thermal hotspots. However, a key limitation of these methods lies in the limited spatial resolution of their sensors: MODIS (1 km), VIIRS (375 m), and Landsat 8 (100 m). Event detection delays are influenced by the time required to transmit the acquired images to ground stations. To mitigate these delays, several end-to-end solutions have been developed, from image acquisition to event detection, to reduce the overall detection latency.

A pioneering end-to-end solution for thermal hotspot detection was demonstrated by the Earth Observing-1 (EO-1) mission (Davies et al.,

2006). As a technology demonstrator, the mission featured autonomous onboard processing and achieved a latency of less than three hours from image acquisition to the ground receipt of detected volcanic activity. The detection algorithm performed a multi-threshold segmentation on 30 m resolution, 5-band SWIR (1.25–2.28 μm) imagery. This system operated on minimally processed data, defined as Level 0.5, which consisted of radiance values with brightness and dark signal corrections applied. This mission is therefore recognized as a pioneer of an end-to-end autonomous system that demonstrated the capability for onboard detection of dynamic events, such as thermal hotspots.

More recent examples of end-to-end solutions include the onboard processing pipeline for low latency fire segmentation proposed by OroraTech (Schöttl et al., 2024) and that proposed for the FireBird mission (Lorenz et al., 2017). OroraTech's approach employs a single-band contextual algorithm (Wooster et al., 2012), utilizing the MIR and TIR bands, with a spatial resolution of 200 m for fire detection. The on-orbit pipeline performs radiometric and geometric corrections onboard, achieving fire detection within less than three minutes. However, a key limitation of this pipeline is the potential for small or early-stage fire events to go undetected due to the 200 m spatial resolution, and the imprecise location of events due to an on-orbit geolocation uncertainty of approximately 1 km. The FireBird mission implemented a threshold solution based on the MIR (3.4-4.2 μm) and LIR (8.5-9.3 μm) spectral bands with a spatial resolution of 178 m. Nevertheless, there are no available data regarding its low-latency capabilities despite its on-orbit deployment.

Other thermal hotspot detection approaches provide conceptual frameworks that address only specific components within the processing pipeline. Zhang et al. (2021) presented an active fire segmentation method for Sentinel-2 using the Red, NIR and SWIR spectral bands. This approach achieved an IoU ∼0.70 with a detection time under 10 min, excluding image pre-processing steps, when tested on a NVIDIA GeForce RTX 2080Ti GPU. Similarly, de Almeida Pereira et al. (2021) proposed a DL-based segmentation solution using a simplified U-Net for the Blue and SWIR bands of Landsat-8, achieving an IoU of 0.814, an F-1 score of 0.897, and an inference time of 25.5 ms per $256 \times 256 \times 3$ patch. Furthermore, Spiller et al. (2022) developed a hyperspectral CNN-based approach for on-orbit wildfire segmentation, implemented on the Movidius Myriad 2, the NVIDIA Jetson NX, and the NVIDIA Jetson Nano. This solution demonstrated power consumptions below 5 W and inference times under 6 ms per pixel. Despite their promising results, these approaches face several limitations. They rely on processed imagery expressed in reflectance values, and their processing times were not evaluated using CubeSat-compatible hardware. These methods only discuss the DL model, overlooking the computationally demanding image pre-processing steps, which impose constraints on both hardware resources and the total detection delay. Moreover, only Spiller et al. (2022) accounts for the inference time on edge computing hardware.

The E2E pipeline, proposed by Meoni et al. (2025), is a more comprehensive conceptual framework for onboard thermal hotspots classification in raw multispectral imagery. The pipeline consists of three main steps: (1) coarse co-registration of the NIR-SWIR spectral bands, (2) tiling of the image into small patches, and (3) DL-based classification of the generated patches. The use of raw imagery reduces the total detection delay and enables the solution to be implemented in less demanding hardware, as it does not require computationally intensive processing steps. This solution was tested on edge computing hardware, specifically the Raspberry Pi 4 and the CogniSAT-XE2 board, achieving a total processing time of 1.8 s and a peak power of 6.4 W. Nevertheless, the main limitation of this method lies in the size of the output classification map, which identifies the occurrence of an event within an approximate area of 25 km^2. Additionally, this concept does not account for the additional delay caused by image acquisition, communications with mass storage, or the post-processing communications required to receive the information on the ground.

A comparative summary of the methods discussed is presented in Table 1. Among them, a key challenge is the trade-off between detection latency and processing complexity. Most methods rely on ground-based analysis, which introduces significant delays as entire images must be downlinked before analysis can begin. While onboard processing can mitigate this latency, most solutions depend on processed imagery, which in turn requires more powerful satellite hardware. Furthermore, numerous solutions employ traditional contextual or threshold-based algorithms, lacking the adaptability of modern AI-based approaches to diverse environmental conditions. Finally, the prevalent use of low spatial resolution imagery inherently limits the detection of small or incipient events, a critical factor for effective early warning.

3. Methodology

The framework of the onboard processing pipeline proposed in this project is depicted in Fig. 1. It consists of four stages: (1) data acquisition and storage, (2) coarse image co-registration, (3) image tiling, and (4) Deep Learning (DL) segmentation inference. Stage 1 encompasses image acquisition, data transfer from sensor to mass storage, and subsequent transfer to the Payload Data Handling Processor (PDHP). The proposed concept excludes this stage from its scope, assuming images are readily available for processing after acquisition, thereby disregarding the expected additional delay. Stage 2 involves a band misalignment correction process performed in the PDHP, ensuring that each pixel corresponds to the same geographical area across the spectral bands. Stage 3 divides the image into smaller tiles to fit the hardware memory constraints on board. Finally, in stage four, the tiles are sequentially transferred from the PDHP to the DL accelerator to perform segmentation and obtain the output segmentation masks.

Its on-board design allows for on-orbit insights extraction, reducing the amount of transmitted data and the delay between image acquisition and data communication with the ground station. Raw multispectral imagery aims to reduce the complexity of the onboard algorithms, reducing the detection delay and enabling the implementation in edge computing satellite hardware, compatible with CubeSats. DL segmentation, pixel-level classification, provides a precise detection solution that addresses the limited generalization capabilities of threshold and contextual methods (Seydi et al., 2022; Hu et al., 2021b). This is obtained through a Fully Convolutional Network (FCN) model specifically designed in this project for on-device implementation, as it requires a fast inference and low power consumption design.

The remainder of this section describes the key elements required for the development of the proposed pipeline. These encompass the creation of the segmentation dataset, the design of the fast-inference DL model, the procedures for model training and validation, and the experimental deployment of the pipeline on representative CubeSat hardware.

3.1. Segmentation dataset

There are no existing datasets for thermal hotspots segmentation in raw multispectral imagery. This is attributed to the lack of available datasets of raw multispectral images, as those are not made available by the data provider. For instance, Sentinel-2 only provides access to level 1C or above, which are already expressed in orthorectified reflectances. However, Meoni et al. (2024) have recently published the Thermal Hotspots in Raw Sentinel-2 data (THRawS) dataset, the first dataset for wildfire and volcanic activity monitoring in raw multispectral imagery. They continued their work in Meoni et al. (2025), using the images to create a classification dataset to detect fires, volcanic eruptions, and non-events. Nevertheless, the precision obtained with classification approaches is limited due to the dependency on the input image size, and it can be further improved by using pixel-level classification. Therefore, this project proposes Segmentation of Thermal Hotspots in

Table 1
Summary of existing thermal hotspot detection methods.

Method	Characteristics
EO-1 (Davies et al., 2006)	• Onboard processing with <3 h from image acquisition to ground receipt of detected thermal hotspots. • Multi-threshold segmentation algorithm on minimally processed multispectral imagery (SWIR) with 30 m spatial resolution. • Pioneer mission for an autonomous onboard system for dynamic events detection in real time.
FIRMS/MIROVA (Davies et al., 2008; Coppola et al., 2016)	• Multi-satellite solution with global NRT coverage (~3 h latency, 1–60 min in US/Canada). • Contextual-based detection algorithms with low spatial resolution (100 m–1 km). • No onboard processing or Inter-Satellite Links (ISL); requires full-image downlink and ground-based processing. • Contextual-based detection algorithm may lack adaptability to diverse environmental conditions.
OroraTech (Schöttl et al., 2024)	• End-to-end onboard solution for fire detection within minutes using thermal bands (MIR and TIR). • Contextual-based detection algorithm with low spatial resolution (200 m) and high geolocation uncertainty (1 km), which limits precise location of small events. • Solution reliant on radiometrically processed imagery, which impacts the overall end-to-end latency.
FireBird (Lorenz et al., 2017)	• Mission with onboard processing capabilities, but no published data on low-latency performance. • Threshold-based detection algorithm with thermal bands (MIR and LIR) and low spatial resolution (178 m).
Zhang et al. (2021)	• High spatial resolution (10–20 m) multispectral AI-based segmentation solution. • Solution focused solely on ground-based detection, not in the end-to-end process, with a detection time lower than 10 min for an entire S-2 scene.
de Almeida Pereira et al. (2021)	• High spatial resolution (30 m) multispectral AI-based segmentation solution compatible with onboard deployment. • Inference time of 25.5 ms per 256 × 256 × 3 image measured on non-representative hardware, without including image processing steps. • Image processing steps are required and not accounted for the potential onboard deployment.
Spiller et al. (2022)	• AI segmentation tested on edge-computing hardware obtaining an inference time of 6 ms/pixel and a power consumption lower than 5 W. • Focuses solely on AI inference, neglecting the image pre-processing steps required for an onboard solution.
Meoni et al. (2025)	• End-to-end methodology using raw multispectral imagery to minimize latency. • Adaptable AI-based classification solution verified on CubeSat-grade hardware. • Methodology obtained an end-to-end time of 1.8 s and a peak power of 6.4 W on CubeSat-grade hardware. • The classification output is an area of 5 × 5 km^2, which limits the precise location of events.

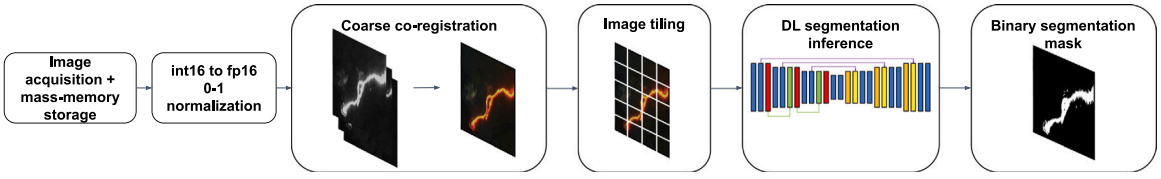


Fig. 1. Diagram of the proposed onboard processing pipeline for thermal hotspots segmentation.

Raw Sentinel-2 (SegTHRawS), the first dataset for thermal hotspots segmentation in raw multispectral satellite imagery.

The raw multispectral satellite imagery is obtained from the THRawS dataset. THRawS contains 146 globally distributed Sentinel-2 raw multispectral images of wildfires and volcanic eruptions between 2016 and 2022 over the highlighted areas in Fig. 2. The images are expressed in raw DN without radiometric processing, and they include the 13 spectral bands of Sentinel-2 ranging from Coastal Aerosol (~0.44 μm) to the Short-Wave InfraRed (SWIR) (~2.2 μm).

The dataset created in this project provides band-aligned images of the entire THrawS dataset. The images have undergone a co-registration process to correct the misalignment between spectral bands. This step is required to align the pixels between bands so that each pixel in the multi-band image contains information on the same area of the Earth's surface. The co-registration process corresponds to the method described in Meoni et al. (2024) with the SuperGlue matching network (Sarlin et al., 2020). In this process, a spectral band acts as a reference band, and the rest are aligned with the reference. The co-registration order for SegTHRawS corresponds to the sequence in which the spectral bands are acquired by the MSI sensor (Meoni et al., 2024). Following this, the co-registered images are normalized between 0–1 to correct the 12-bit radiometric resolution and cropped into 256 × 256 patches with 25% overlap in every direction to include potential events in the edges.

The segmentation masks are obtained through a strict majority voting approach that performs automatic labeling to avoid manual annotation based on visual inspection. The mask generation process is based on the majority voting principle described in de Almeida Pereira et al. (2021). This process combines the outputs of several state-of-the-art detection methods and compares them to generate a segmentation mask based on voting. For instance, the article employed three state-of-the-art thermal hotspots segmentation threshold methods to create a binary segmentation mask with a positive class on those pixels which at least two methods, the majority, agreed. For this project, a stricter majority voting with five methods is proposed, as there are no existing methods for thermal hotspots detection in raw multispectral imagery. The five methods correspond to four extracted from the literature (Schroeder et al., 2016, Murphy et al., 2016, Kumar and Roy, 2018, and Massimetti et al., 2020), and the novel Castro-Traba conditions specifically designed in this project for raw multispectral imagery. This new method was incorporated to minimize false detections and eliminate reliance on subjective visual inspection.

The segmentation masks of the strict majority voting follow a weakly-supervised segmentation approach to make the annotations more robust (Schmitt et al., 2020). Weak supervision involves labeling uncertain pixels separately, excluding them during model training. With this solution, a pixel is categorized as an "event" if at least four out of five methods classify it as such, as a "potential event" if only two or three out of five methods agree, and as "notevent" otherwise. This leads to masks with three possible classes (−1,0,1) for "notevent", "potential_event" and "event", respectively.

Castro-Traba conditions correspond to the fifth method used for automatic labeling. These conditions take inspiration from the four previous methods and add new conditions directly derived from the raw

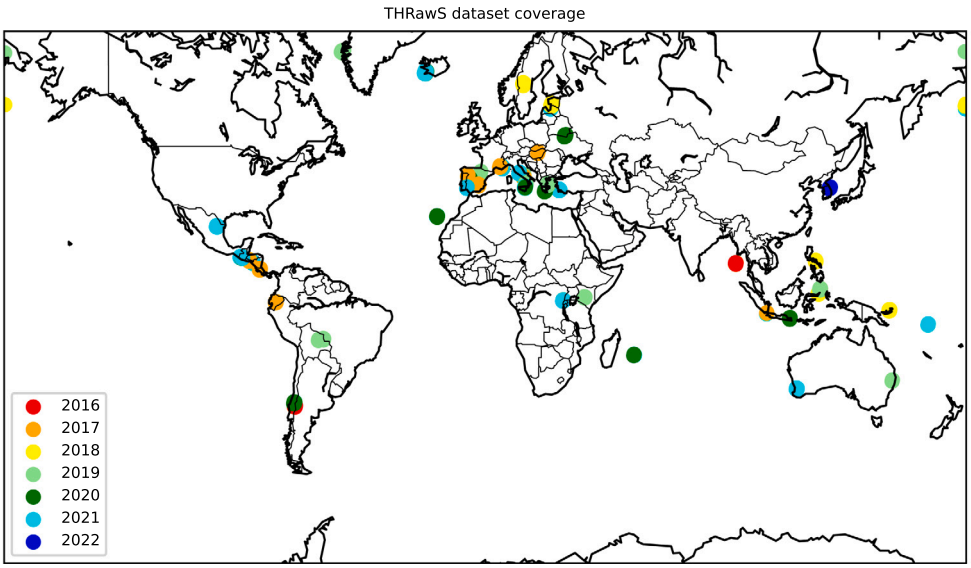


Fig. 2. THRawS dataset global coverage.

imagery. The thresholds were determined through visual inspection of the images with events in the THRawS dataset. This involved extracting pixel values in the NIR-SWIR bands, expressed in int8 format, and identifying patterns associated with visually identified event pixels. In the equations, DN_i is the raw digital number (DN) of channel i, $DN_{N,i}$ the raw digital number of a 30 × 30 neighborhood of channel i, $R^{DN}_{i,j}$ is the ratio between raw DNs in channels i and j, and $R^{DN}_{N_{i,j}}$ is the ratio between raw DNs of the pixels in a 30 × 30 neighborhood of channels i and j.

The first set of conditions depicted in Eq. (1) is the same as the first conditions from Murphy et al. (2016). The next condition, Eq. (2), aims to reduce the number of false detections by adding stricter thresholds to the SWIR2 band (B12) and the NIR2 band (B8A). Spectral bands near higher wavelengths are more sensitive to thermal hotspots (Massimetti et al., 2020), and SWIR bands have a peak of radiation for wildfires according to Wien's displacement law (Wooster et al., 2012). This effect is seen in the raw images with high DN values for the SWIR2 band and low DN values for the NIR2 band. This condition performs a logical AND operation with Eq. (1).

$$\left(R^{DN}_{12,11} \geq 1.4\right) \ \& \ \left(R^{DN}_{12,8A} \geq 1.4\right) \ \& \left(DN_{12} > 0.15\right) \tag{1}$$

$$\left(DN_{12} > \frac{100}{255}\right) \ \& \ \left(DN_{8A} < \frac{50}{225}\right) \tag{2}$$

Eq. (3) is designed for thermal events with similarly high values for both SWIR bands and relatively low values for the NIR2 band. This condition aims to detect the core of thermal hotspots and performs a logical OR operation with the output mask from Eq. (2). Eq. (4) detects the hotter part of the thermal hotspots, where both SWIR bands present the highest values. It performs a logical OR operation with the output mask from Eq. (3).

$$\left[\left(DN_{12} > \frac{150}{255}\right) \text{ or } \left(DN_{11} > \frac{150}{255}\right)\right] \& \ \left(DN_{8A} < \frac{60}{255}\right) \tag{3}$$

$$\left(DN_{12} > \frac{180}{255}\right) \ \& \ \left(DN_{11} \geq \frac{180}{255}\right) \tag{4}$$

After the threshold conditions, Eq. (5), a contextual condition, is applied to detect potential neighboring pixels affected by thermal hotspots. The neighborhood selected is 30 × 30 centered for each of the previously detected pixels, and those neighboring pixels that satisfy Eq. (5) are classified as thermal hotspots. This condition performs a logical OR with the output mask from Eq. (4). As a final postprocessing step, a non-event pixel is classified as a thermal hotspot if it is completely surrounded by thermal hotspot pixels.

$$\left(R^{DN}_{N_{12,11}} \geq 4\right) \ \& \ \left(DN_{N,8A} < 0.1\right) \tag{5}$$

These conditions were verified by visual inspection, following the same approach used in similar studies (Shirvani et al., 2023; Zhang et al., 2021; Hu et al., 2021a; Massimetti et al., 2020; de Almeida Pereira et al., 2021; Schroeder et al., 2016; Murphy et al., 2016). Fig. 3 shows the detection performance compared to the remaining methods used in the majority voting process, and the majority voting 4 (where 4/5 methods classify the pixel as an event). The performance is comparable to that of the rest of the methods for raw imagery. Based on these results, these conditions were included as an additional method in the segmentation mask creation process.

Table 2
Pixel proportion statistics of the event images in SegTHRawS.

	Event	Potential event	Notevent
Total # of pixels	$1.367 \cdot 10^5$	$1.261 \cdot 10^5$	$1.077 \cdot 10^8$
Mean # of pixels	83.01	76.58	$6.538 \cdot 10^4$
Std # of pixels	362.9	361.5	699.5
Total %	0.1270	0.1170	99.76

The SegTHRawS dataset comprises co-registered raw multispectral images stored in Planar binary format (.bin), totaling 17,817 images categorized into 824 events, 14,099 non-events, and 2894 potential events, with a combined data size of 50.4 GB. Each image has a shape of 256 × 256 pixels and is normalized within the range of 0–1, encompassing the 13 spectral channels of Sentinel-2. Segmentation masks, generated through the weakly-supervised approach, are available exclusively for event and potential event images. The dataset exhibits a substantial imbalance, with non-event images considerably outnumbering event images. This imbalance extends to the pixel level within the segmentation masks, as illustrated in Table 2 for the masks of the event images. This imbalance poses challenges for model training, potentially biasing predictions towards the majority class, thus limiting model detection performance (Cariello et al., 2023). This limitation is examined and mitigated using the strategy discussed in Section 3.3.

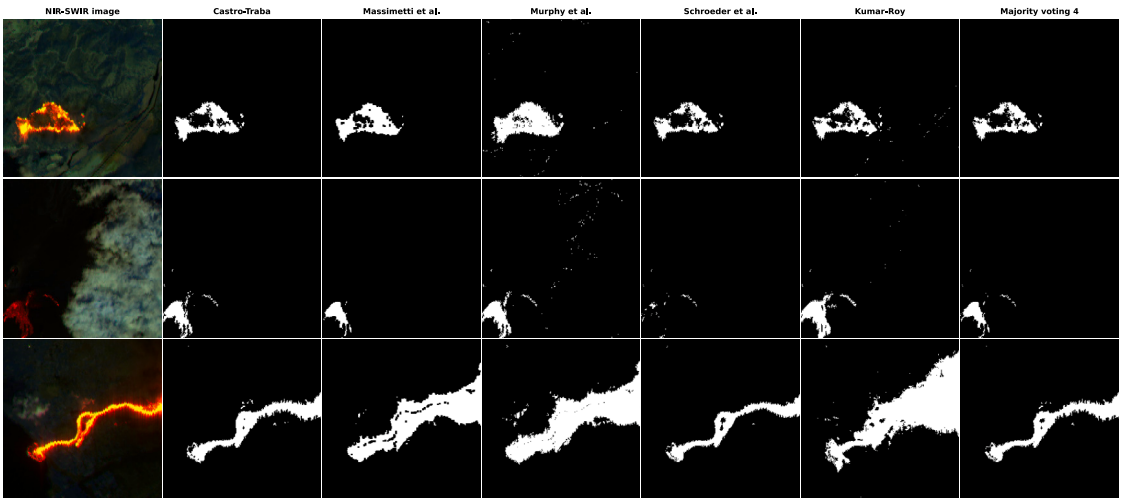


Fig. 3. Detection performance of the five contextual threshold approaches and the majority voting 4 for three cases, displayed in rows. (a) Fire event detected in France. (b) Volcanic event detected in Mount Etna, Italy (c) Volcanic event detected in La Palma, Spain.

3.2. Onboard-tailored DL segmentation models

The requirement for onboard processing with low-latency significantly influences the selection of the segmentation architecture, requiring a balance between inference time, complexity, and detection performance. Several architectures combine a segmentation network with an efficient backbone focused on on-device fast inference. Nevertheless, smaller and simpler models, derived from existing architectures, can potentially yield faster inference times while maintaining comparable detection performance. Hardware considerations are also crucial in architecture selection. This study employs the same hardware platform as that used in Meoni et al. (2025), which demonstrated the feasibility of in-orbit processing. Building upon the original U-Net architecture (Ronneberger et al., 2015), optimized for the Myriad, this work presents a modified version for fast inference and onboard implementation. These modifications aim to reduce model complexity and potentially improve inference times without sacrificing detection performance. The specific modifications are as follows:

- **Increase bottleneck size**: Remove the deepest level of the original U-Net architecture. U-Net was designed for 572 × 572 images, whereas this project uses 256 × 256 images, less than half the original. Since the image size is halved at each level, removing the deepest level results in a network with an image resizing ratio between the input and bottleneck comparable to that of UNet.
- **Reduce the number of filters per level**: Reducing the number of filters per level significantly decreases the number of model parameters, reducing training and inference times at the potential cost of detection performance.
- **Encoder residual connections**: Adding residual connections in the encoder to maintain the detection capability of the less complex model. Residual blocks improve the detection capability of a model, as they combine information from layers of different levels, and address the problem of vanishing gradients present in CNNs.

These modifications can be implemented in any Fully Convolutional Network (FCN) for on-device implementation. In this project, U-Net, Attention U-Net, and U-Net 3+ were considered to evaluate these modifications concerning detection performance and inference time. For Attention U-Net and U-Net 3+, only the first and second modifications were considered. The modified models are named by adding "-SX" to the original model name, where X is an integer from 0 and 5 indicating the model size based on the number of filters per level. The visual representation of these modifications for the ResUNet-S2 are depicted in Fig. 4

3.3. Model training

Only segmentation models designed for on-device implementation are considered for training. These models include those outlined in Section 3.2, as well as U-Net, U-Net++ and DeepLabV3+ architectures, each paired with MobileOne-S0, EfficientNet-B0, and MobileNetV2 encoders.

The SegTHRawS dataset is unsuitable for training a segmentation model due to a pronounced class imbalance between images containing thermal events and those without. The class imbalance problem was discussed in Section 3.1, both at the image and pixel levels. Segmentation models require a balanced training dataset between positive (event) and negative (non-event) classes to mitigate the risk of bias towards the majority class prediction (Cariello et al., 2023), even with the application of specific loss functions designed to address class imbalance.

To address this issue, two strategies were considered: (1) reducing the dataset to achieve a 50/50 balance between event and non-event images (Cariello et al., 2023), and (2) augmenting event images until their quantity matches that of no-event images (Del Rosso et al., 2021). However, the second strategy is more effective in datasets with a lower degree of imbalance, as excessive augmentation can lead to augmented images that are highly similar to the originals, increasing the risk of model overfitting. Therefore, we propose constructing a reduced training dataset derived from SegThRawS, maintaining a balanced 50/50 proportion of event and non-event images to mitigate class imbalance. This dataset employs a geographical split strategy, whereby specific geographical areas are exclusively reserved for testing. This strategy provides a rigorous evaluation of the model's detection performance on scenes not considered during training and validation, thereby assessing its generalization capability on unseen data. In Fig. 5, the process for generating the 50/50 balanced training dataset with a geographical split is depicted.

The training dataset contains images and weakly labeled masks in Planar binary format, with shapes [1,256,256,3] and [1,256,256,1], respectively. This dataset comprises 1648 images, with 1320 images allocated for training, 164 for validation, and 164 for testing, with a total size of 1.7 GB.

Despite utilizing the 50/50 balanced dataset, a notable class imbalance persists between event and non-event pixels. To address this challenge, the Focal Loss is employed, which was specifically designed to mitigate class imbalance and prioritize difficult-to-classify instances (Lin, 2017). Model training was conducted using PyTorch

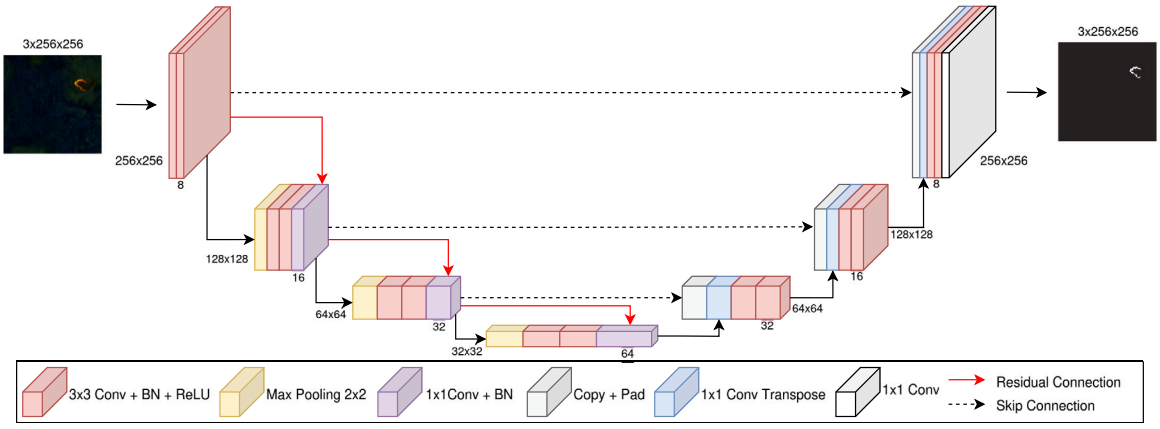


Fig. 4. Visual representation of the ResUNet-S2 network proposed in this project.

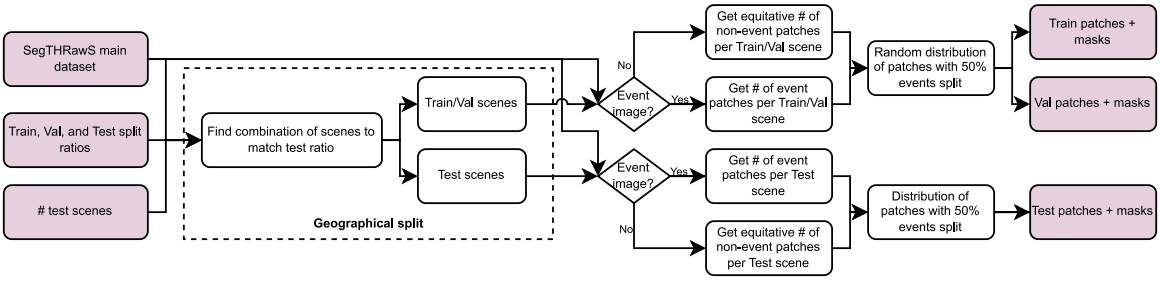


Fig. 5. Diagram of the process for creating the 50/50 balanced training dataset with a geographical split.

Lightning.[1] The Focal Loss function was applied with an α parameter of 0.5, representing the class weighting factor, and a γ value of 2 to emphasize hard-to-segment cases. The learning rate was determined using Pytorch Lightning's learning rate finder for most models (Smith, 2017), except for the modified U-Net models, which employed a fixed learning rate of $3e^{-4}$. The "ReduceLROnPlateau" scheduler was also applied to adjust the learning rate during training. Each model underwent training over five random seeds (0,1,2,42,73) using mixed precision, a batch size of 24, weak labeling, flip and rotation non-destructive image augmentations, no pre-trained weights, and using the Adaptive Moment Estimation (Adam) optimizer. The training procedures took place on an NVIDIA A30 with 12 GB of Video Random-Access Memory (VRAM) and an NVIDIA GeForce RTX 4060 Laptop with 8 GB VRAM. The code for the training workflow is publicly accessible on the designated GitHub repository, referenced in Data availability.

3.3.1. Metrics

Due to the significant class imbalance in the dataset, conventional metrics such as accuracy and precision do not adequately reflect the model's detection performance. For instance, predicting all pixels in the reduced training dataset as non-events would yield an accuracy of approximately 99%, which is misleading. Therefore, segmentation performance is evaluated using the IoU, also referred to as the Jaccard index, and the F-1 score (Taha and Hanbury, 2015). To ensure robustness, these metrics are calculated independently for each class, thereby avoiding the occurrence of Not a Number (NaN) values in cases where some masks may be empty.

For instance, if the model predicts all pixels in the small training dataset as "notevent" the accuracy would be higher than 99%, which is misleading. To evaluate segmentation performance, the key metrics correspond to the F-score and the IoU, also known as the Jaccard index (Taha and Hanbury, 2015). Both metrics are calculated separately for each class to avoid NaN values in empty mask images.

3.4. DL model validation in an external dataset

The final segmentation model was tested on an independent external segmentation dataset for thermal hotspot detection to assess its generalization capability and rule out overfitting. This evaluation strategy, combined with the geographical split in the training dataset, strengthens the evidence for the model's robust segmentation performance.

The external dataset utilized for this purpose corresponds to the de Almeida Pereira et al. (2021) dataset, focused on thermal hotspots segmentation in Landsat-8 L1C reflectance-based imagery. This dataset was selected for its widespread availability, global coverage of thermal events, and because the authors used a lightweight U-Net architecture that can be implemented on-board. Given the extensive size of the dataset, only 20% of the total images were allocated for training and validation, while maintaining consistency with the same manually annotated testing set. The training images were derived from Oceania, North America, Europe, and Eastern Asia, and were gathered using a 50/50 balanced split of events and non-events for each continent. This approach follows the procedure illustrated in Fig. 5. Two test configurations are considered for evaluation: (1) zero-shot segmentation, and (2) retraining the model on the new dataset. The retraining procedure involves leveraging the model with pre-trained weights, obtained from training on the training dataset of SegTHRawS, as a foundational starting point for further training on the new dataset.

[1] https://lightning.ai/pytorch-lightning

3.5. Experiments

The proposed payload processing pipeline is validated on edge computing CubeSat hardware to assess its expected on-orbit performance. This experimental evaluation measures both the total processing time per pipeline execution and the energy and power consumption profiles of various onboard models tested within the operational hardware environment. The models included in the evaluation are the U-Net+MobileNetV2 model from the literature, the ResUNet-S models introduced in this work, and the Att-Unet-S0 and Att-Unet-S4 models, also proposed in this study. The objective is to quantify the inference time of the proposed models and compare their performance with that of a state-of-the-art on-device model.

The pipeline operates on a 12-bit binary input image captured in the NIR-SWIR spectral bands, with a resolution of $1152 \times 1296 \times 3$ pixels. One execution of the pipeline comprises three primary steps: (1) coarse co-registration of the three spectral bands, (2) tiling of the co-registered image into 20 non-overlapping patches, each of size $256 \times 256 \times 3$ patches, and (3) segmentation inference performed on the 20 patches.

The selected hardware configuration consists of the Raspberry Pi 3B+ as the PDHP and the Intel Movidius Myriad X VPU serving as an AI accelerator. The Raspberry Pi 3B+ handles the pre- and post-processing stages of the pipeline and was chosen due to its easy availability and its ARM processor, which makes it a Commercially Off-The-Shelf (COTS) PDHP suitable for low-power space applications. This processor has been considered in several SmallSats and CubeSats designs (Yang et al., 2017; Mohd-Isah et al., 2020). The Myriad X, responsible for the AI segmentation inference within the pipeline, leverages two LEON4 cores as Central Processing Units, 4 Gbit of Dynamic Random Access Memory, and the Neural Compute Engine, a dedicated hardware accelerator designed for on-device, high-speed, and low-power AI inference (Dunkel et al., 2022). It represents the next generation of the Myriad 2, previously deployed in the Φ-sat 1 (Giuffrida et al., 2021) and the Φ-sat 2 (Longepé et al., 2023) missions. In this study, the Myriad X is presented within the CogniSAT-XE2[2] board developed by Ubotica Technologies. This board is built around the Myriad X VPU to provide a low-power high-speed solution for AI inference onboard SmallSats and CubeSats. It has been deployed and verified in the space environment in the CogniSAT-6 mission (Rijlaarsdam et al., 2025).

The experimental setup, depicted in Fig. 6, illustrates the configuration, comprising the Siglent SPD3000X power supply, the Raspberry Pi 3B+, and the CogniSAT-XE2, arranged from top to bottom.

4. Results

4.1. Model validation in an external dataset

The model selected for evaluation in the external dataset was the ResUNet-S2 model proposed in this project. Table 3 reports the segmentation results of the ResUNet-S2 model evaluated on the external dataset from de Almeida Pereira et al. (2021). The retrained ResUNet-S2 model, proposed in this study, achieves an IoU of 0.835 and an F-1 score of 0.909, outperforming the U-Net-Light (3c) used in the original article. It should be noted that the ResUnet-S2 model was trained using only 20% of the total available training data, employing the 50/50 balanced dataset approach outlined in Section 3.3. Fig. 7 illustrates a comparative analysis of detection performance between the proposed segmentation model and the network described in the original study across three distinct cases. These results demonstrate the detection and generalization capabilities of the ResUNet-S2 model in a different dataset, which provide evidence on the absence of overfitting of the model in the SegTHRawS dataset.

[2] https://ubotica.com/ubotica-cognisat-xe2/

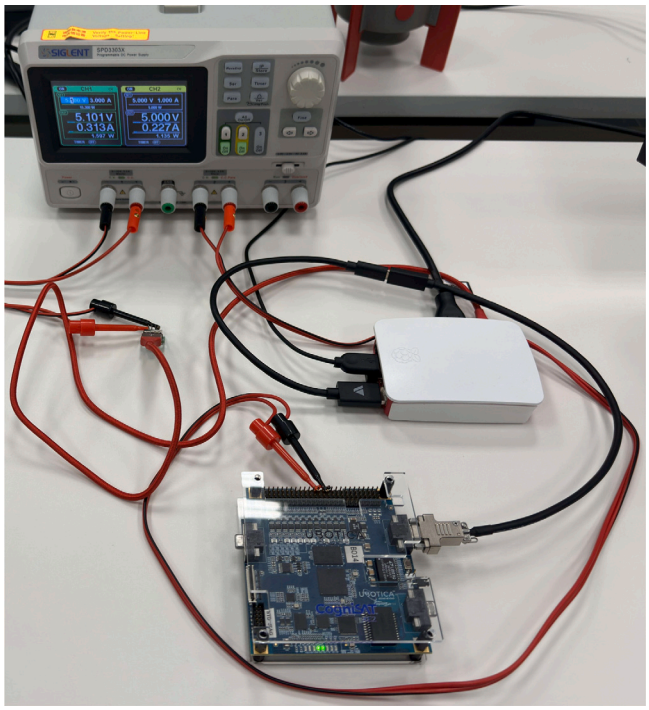


Fig. 6. Experimental setup of the Raspberry Pi 3B+ and CogniSAT-XE2 board experiment.

Table 3
ResUNet-S2 detection performance in de Almeida Pereira et al. (2021) dataset.

Method	IoU	F-1
Zero-shot ResUNet-S2	0.6944	0.816
Retrained ResUNet-S2	**0.835**	**0.909**
U-Net-Light (3c) (de Almeida Pereira et al., 2021)	0.814	0.897

4.2. Experiment results

Table 4 presents the results of a single pipeline execution on the Raspberry PI 3B+ and the CogniSAT-XE2 for the selected models. This table includes the model size and the inference time measured in the CogniSAT-XE2 board per $256 \times 256 \times 3$ image patch for each tested model. The total processing time and energy correspond to the complete execution of the pipeline, from the input image to the output segmentation mask. The total energy is represented by two values: (1) the energy calculated by multiplying the processing time and the peak power of each pipeline stage, and (2) the measured energy, which accounts for the additional energy consumption associated with hardware initialization. Additionally, Fig. 8 illustrates the power profile for a single pipeline execution using the ResUNet-S2 model.

Fig. 9 displays the output segmentation masks produced by the ResUNet-S2 in the CogniSAT-XE2 experiment for 5 images from the test dataset. The second and fourth rows exemplify the model's performance under nominal conditions, showing the accurately segmented, clearly visible volcanic eruptions of the Mayon and the Piton de la Fournaise volcanoes, respectively.

In contrast, the first and third rows demonstrate cases where the model's performance is degraded, but still relatively high IoU, by the presence of clouds near or over the event. The first row, however, also highlights the importance of weak labeling, as the model segments pixels that were only considered potential events in the initial dataset. The third row shows an instance where the model fails to properly segment low-intensity events obscured by clouds, generating larger segmentation regions than expected based on visual verification. This

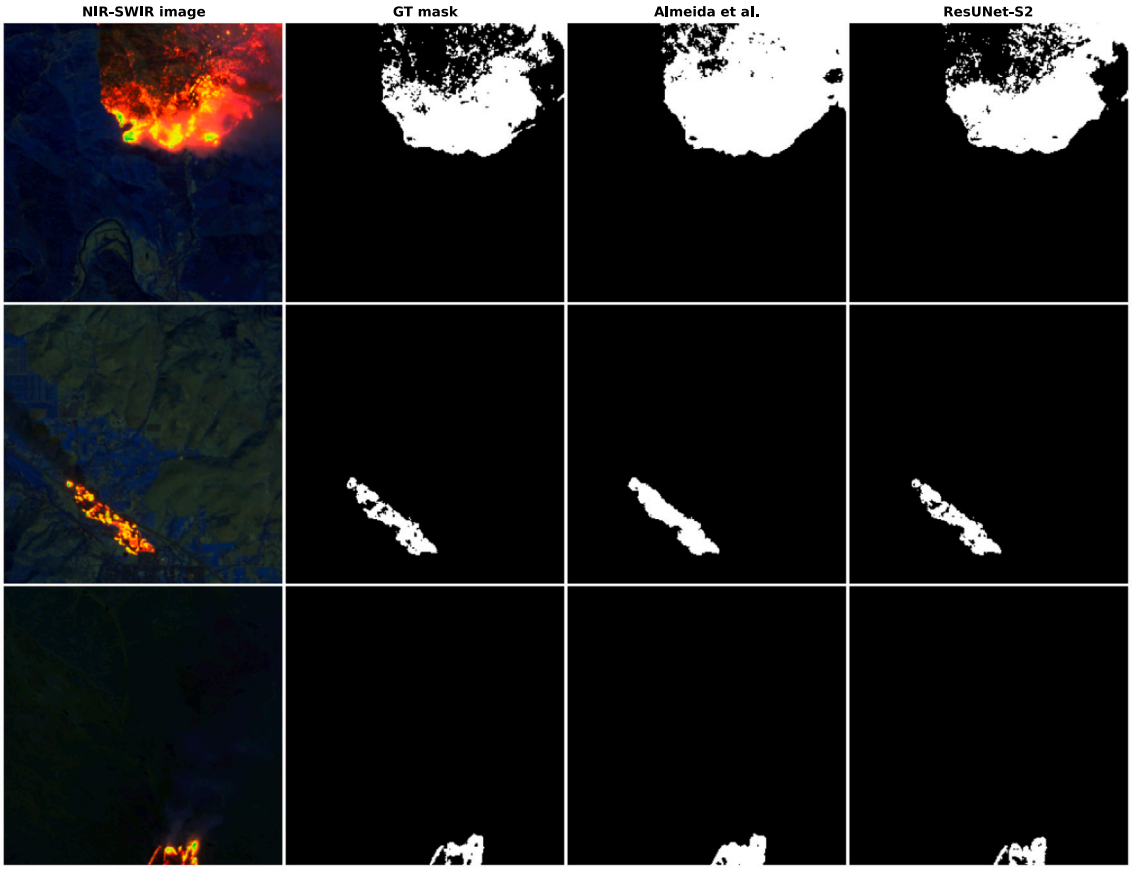


Fig. 7. Comparative analysis of the segmentation outputs in the de Almeida Pereira et al. (2021) dataset of the ResUNet-S2 model and the network of the original article.

Table 4
Detection and time performance of the tested models in the Raspberry Pi 3B+ and CogniSAT-XE2.

Model	Model size [MB]	IoU	F-1	Inference time [ms]	Time [s]	Total energy [J]	Peak power [W]
Unet MobileNetV2	25.43	0.977	0.988	69.89	1.913	7.954	4.470
ResUNet-S0	0.031	0.945	0.974	**28.75**	**1.022**	**3.946**	4.190
ResUNet-S1	0.121	0.938	0.967	37.47	1.191	4.691	4.265
ResUNet-S2	0.477	**0.988**	0.986	50.28	1.452	5.833	4.062
ResUNet-S3	1.891	0.973	0.986	76.24	1.976	8.140	4.430
ResUNet-S4	7.533	0.981	**0.991**	131.5	3.085	11.64	**4.060**
Att-UNet-S0	0.037	0.945	0.971	42.32	1.283	4.953	4.175
Att-UNet-S4	8.282	0.981	**0.991**	153.2	3.517	16.38	4.960

discrepancy could be attributed to two potential causes. First, radiation from the event could be reflecting off the clouds, creating intensity values that the model wrongly segments as low-intensity events. Alternatively, the issue may come from the dataset creation process itself, where the majority-voting strategy could have mislabeled pixels surrounding cloudy events, thereby causing the model to incorrectly learn to segment these areas as part of the event.

Finally, the fifth row illustrates the model's performance on a small, subtle volcanic event at Piton de la Fournaise. This case highlights the importance of the proposed solution for detecting small-scale events (<50 m), which higher-resolution approaches might otherwise fail to detect.

5. Discussion

This study introduces the concept of a novel onboard payload processing pipeline for low-latency segmentation of thermal hotspots using raw multispectral imagery. This solution aims to reduce the time delay in conventional approaches, where "real-time" performance is based solely on post-downlink processing, thus neglecting the latency associated with data transmission and the advantages of onboard processing. The onboard capabilities of the pipeline enable insights to be extracted on-orbit, which reduces the data to be transmitted, and consequently, the existing delay until the detected event is seen on the ground. As can be seen from Table 4, one execution of the pipeline from the input image to the output segmentation masks takes 1.452 s. This time is lower than the 3.6 s that Sentinel-2 requires to capture a granule (European Space Agency - ESA, 2018), demonstrating that this solution can potentially detect thermal events in all the acquired images.

For comparison, FIRMS (Davies et al., 2008) provides 1–60 min detections, after a satellite pass over a ground station, in specific areas of North America and ~3 h globally, with 375–1 km spatial resolution. On the other hand, the proposed onboard pipeline detects the thermal event directly onboard in less than 2 s after image acquisition, with a spatial resolution of 20 m. While the proposed concept focuses on the core processing pipeline, it is crucial to note that additional delays arise from image readout, mass-memory communications, and post-processing communications. This solution can be combined with an

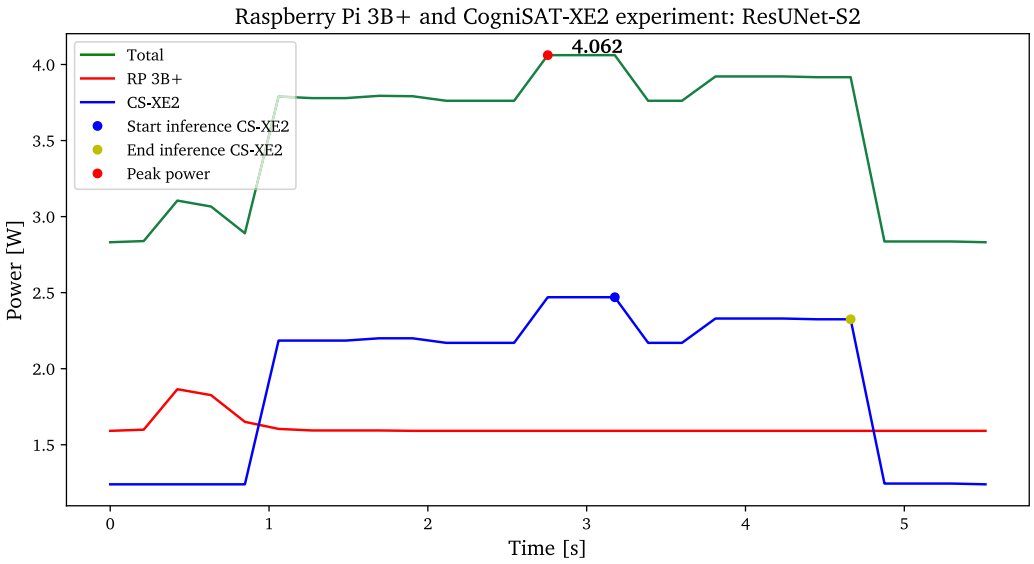


Fig. 8. Power measurement of the ResUNet-S2 model for the Raspberry Pi 3B+ and CogniSAT-XE2 experiment.

ISL to directly communicate the location of events independently of the satellite's orbital position, removing the need to wait for a ground station pass.

The pipeline was tested on CubeSat-compatible edge computing hardware to demonstrate its suitability for energy-constrained environments. One execution consumes less than 16 J, and obtains a peak power of 4.062 W, an IoU of 0.988, and an F-1 score of 0.986. These results demonstrate the energy efficiency of the pipeline while maintaining state-of-the-art detection performance. The detection performance improvement in the CogniSAT-XE2 compared with PyTorch can be attributed to the quantization between float32 and float16 which may have reduced the existing overfitting in the model, or to some batch normalization layers that were modified during the ONNX conversion of the model.

The results from Table 4 provide interesting insights into the relationship between inference time and model size of the proposed models in this project. From the ResUNet models, it can be inferred that for the same type of architecture, the smaller the model the lower the inference time. Nevertheless, when changing the network more factors come into action, as can be seen from the higher inference time of Att-UNet-S0 compared to ResUNet-S0 despite being smaller. The increase in inference time is attributed to the complexity of the attention layers, which corroborates the results from the original article of the Attention-UNet (Oktay, 2018).

Fig. 10 illustrates three 2D analyses of the false detections of the ResUNet-S2 model concerning the intensity values of the NIR-SWIR images in the training dataset. False detections occur mainly when the three spectral bands have relatively similar values, with the NIR band the lowest among the three. This can be attributed to the fact that thermal hotspots are typically associated with high contrast between NIR and SWIR bands. False detection in low-intensity values is expected, as they are not typically associated with thermal hotspots.

The potential origin of these false detections might be attributed to the quality of the segmentation masks. The masks were created through an automatic majority voting labeling approach with four reflectance-based methods without perfect detection performance. As a result, some pixels may have been misclassified as events. These pixels may include the cases of reflective clouds and moderate B11 and B12 values, which may be detrimental to model training.

Based on the results of the detection performance depicted in Table 4, one can assume that the ResUNet-S2 model may suffer from overfitting. However, a geographical split was used during model training to improve the model's generalization capability in unseen data. In addition to this, the ResUnet-S2 model has been validated in the Almeida et al. (de Almeida Pereira et al., 2021) thermal hotspots segmentation dataset, to ensure that the model detection performance was not caused by overfitting. Due to its large size, only 20% of the total images were used for training and validation, while maintaining the same manually annotated testing dataset. Through this approach, the re-trained ResUnet-S2 model has obtained detection metrics that are superior to the model proposed in the original article. The zero-shot segmentation performance and the geographical split used model training in the SegTHRawS dataset indicate the good generalization capabilities of the trained ResUNet-S2 model. Moreover, the detection performance after retraining in the new dataset proves the detection capabilities of the modified models generated in this project. These three reasons provide compelling evidence to infer that the model has not overfitted.

6. Conclusion

This study has proposed the first payload processing pipeline for onboard segmentation of thermal hotspots using raw multispectral satellite imagery. The pipeline was tested using edge computing CubeSat hardware, the Raspberry PI 3B+ and the CogniSAT-XE2 board, to demonstrate its low latency and low power consumption performance. Segmentation is performed through a modified U-Net, ResUNet-S2, proposed in this project for on-device fast inference applications. The model was trained using the SegTHRawS dataset, the first segmentation dataset for thermal hotspot detection in raw multispectral satellite imagery. The segmentation masks were generated following a weak labeling and strict majority voting process among four contextual methods from the literature and one contextual method proposed in this study. The trained ResUNet-S2 implemented in the CogniSAT-XE2 obtained an IoU of 0.988 and an F-1 score of 0.986 in the SegTHRawS dataset. To demonstrate the absence of overfitting and the generalization capabilities of ResUNet-S2, this model was validated in an external dataset for thermal hotspot segmentation with a different satellite sensor, obtaining better detection performance than the original article. A limitation of this project is the majority voting method used for the mask generation process of SegTHRawS, which is mostly based on algorithms that rely on TOA reflectance and not on raw DNs. The Castro-Traba conditions proposed, weakly labeled masks, and the strict

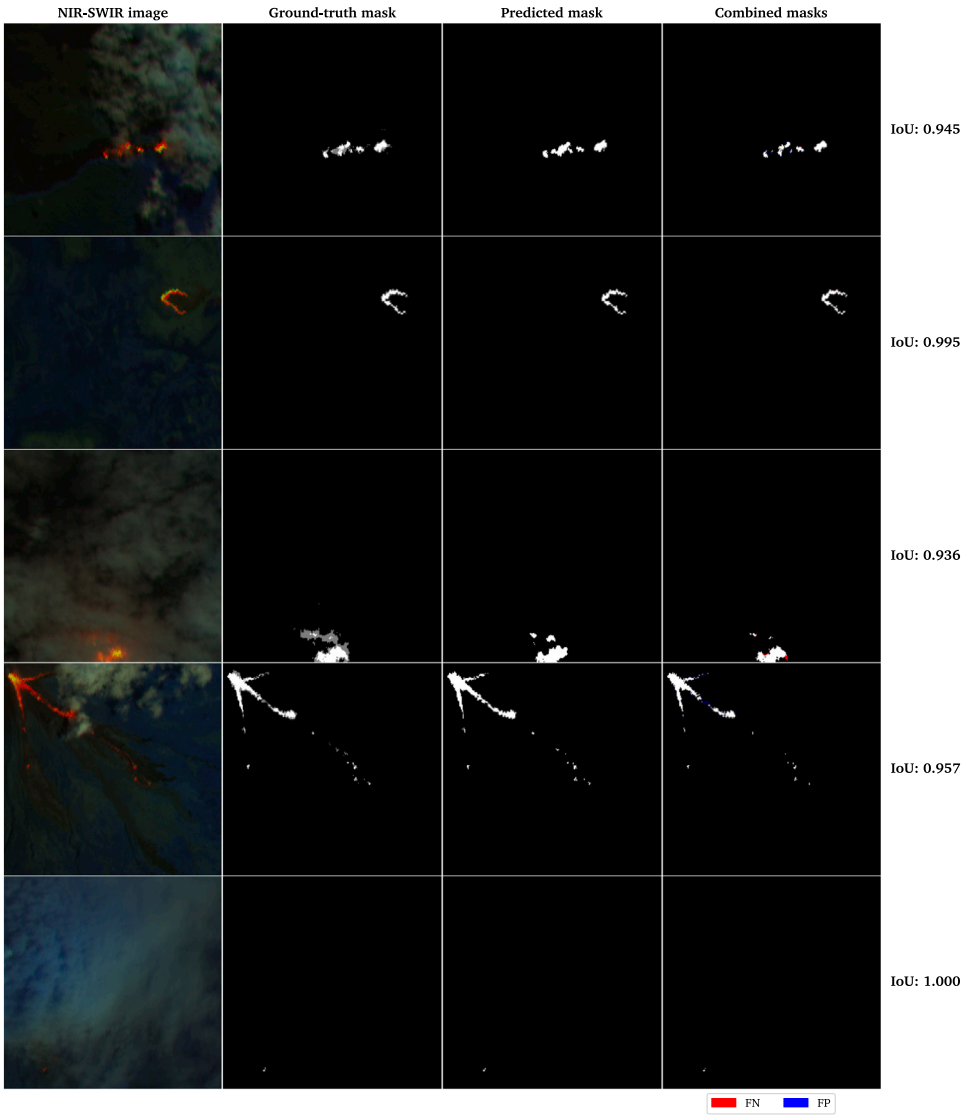


Fig. 9. Segmentation performance of the ResUNet-2 model in the CogniSAT-XE2 for five particular cases of the test dataset. The columns present the NIR-SWIR image, the ground truth mask, the predicted mask obtained in the CogniSAT-XE2, and the combined mask between ground-truth and prediction, where the FP and FN are highlighted, in blue and red, respectively.

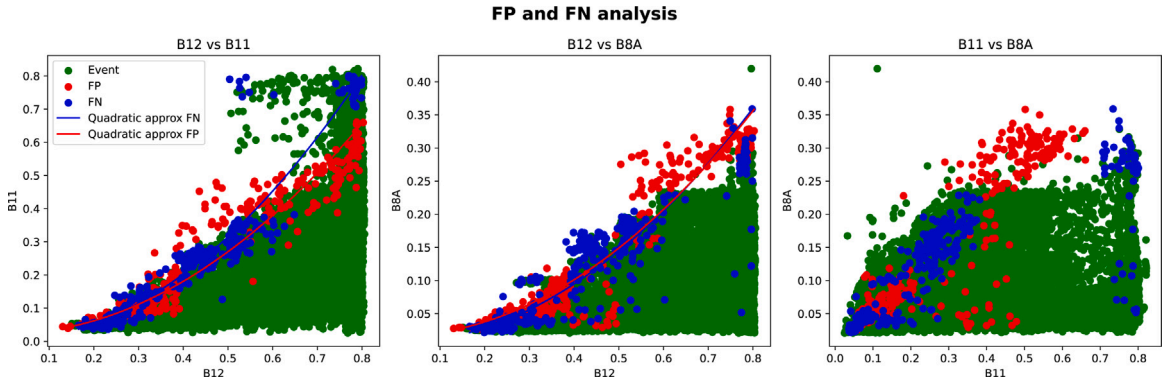


Fig. 10. FP and FN 2D analyses of the NIR-SWIR band values. From left to right: B12 vs B11, B12 vs B8 A, and B11 vs B8A.

majority voting with 4/5 methods tried to address this limitation. Furthermore, the proposed concept works under two assumptions: (1) it assumes the immediate availability of imagery for processing after acquisition, neglecting potential delays associated with data transfer from the sensor to mass storage, and (2) it does not account for the communication time required to downlink the segmentation results to ground stations. Further work will study the use of different spectral band combinations to detect thermal events, the addition of new images

into the SegTHRawS dataset, and the development of a sensor-agnostic thermal hotspots segmentation pipeline.

CRediT authorship contribution statement

Cristopher Castro Traba: Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **David Rijlaarsdam:** Writing – review & editing, Supervision, Resources, Conceptualization. **Jian Guo:** Writing – review & editing, Supervision, Resources, Conceptualization. **Roberto Del Prete:** Writing – review & editing, Supervision, Resources, Conceptualization. **Gabriele Meoni:** Writing – review & editing, Supervision, Resources, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The source code for this research is publicly available at https://github.com/Ubotica/SegTHRawS_ext and includes instructions to generate the training dataset, as well as to train and test the models. The corresponding dataset, which contains all satellite imagery and segmentation masks used in this study, is publicly available at Zenodo with the following link: https://zenodo.org/records/14741990.

References

Bowman, D.M., Williamson, G.J., Abatzoglou, J.T., Kolden, C.A., Cochrane, M.A., Smith, A.M., 2017. Human exposure and sensitivity to globally extreme wildfire events. Nat. Ecol. Evol. 1 (3), 0058. http://dx.doi.org/10.1038/s41559-016-0058.

Bretschneider, T., Vladimirova, T., Yuhaniz, S., 2005. Image processing capabilities on-board micro-satellites for disaster monitoring. In: Proceedings of Second Asian Space Conference, ASC.

Cariello, S., Corradino, C., Torrisi, F., Del Negro, C., 2023. Cascading machine learning to monitor volcanic thermal activity using orbital infrared data: From detection to quantitative evaluation. Remote. Sens. 16 (1), 171. http://dx.doi.org/10.3390/rs16010171.

Chen, T., Liu, Y., 2021. A quick band-to-band mis-registration detection method for sentinel-2 MSI images. Remote. Sens. 13 (17), 3351. http://dx.doi.org/10.3390/rs13173351.

Coppola, D., Laiolo, M., Cigolini, C., Donne, D.D., Ripepe, M., 2016. Enhanced volcanic hot-spot detection using MODIS IR data: results from the MIROVA system. Geol. Soc. Lond. Spec. Publ. 426 (1), 181–205. http://dx.doi.org/10.1144/SP426.5.

Davies, A., Chien, S., Baker, V., Doggett, T., Dohm, J., Greeley, R., Ip, F., Castan, R., Cichy, B., Rabideau, G., et al., 2006. Monitoring active volcanism with the autonomous sciencecraft experiment on EO-1. Remote Sens. Environ. 101 (4), 427–446. http://dx.doi.org/10.1016/j.rse.2005.08.007.

Davies, D.K., Ilavajhala, S., Wong, M.M., Justice, C.O., 2008. Fire information for resource management system: archiving and distributing MODIS active fire data. IEEE Trans. Geosci. Remote Sens. 47 (1), 72–79. http://dx.doi.org/10.1109/TGRS.2008.2002076.

de Almeida Pereira, G.H., Fusioka, A.M., Nassu, B.T., Minetto, R., 2021. Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study. ISPRS J. Photogramm. Remote Sens. 178, 171–186. http://dx.doi.org/10.1016/j.isprsjprs.2021.06.002.

Del Rosso, M.P., Sebastianelli, A., Spiller, D., Mathieu, P.P., Ullo, S.L., 2021. On-board volcanic eruption detection through CNNs and satellite multispectral imagery. Remote. Sens. 13 (17), 3479. http://dx.doi.org/10.3390/rs13173479.

Dunkel, E., Swope, J., Towfic, Z., Chien, S., Russell, D., Sauvageau, J., Sheldon, D., Romero-Cañas, J., Espinosa-Aranda, J.L., Buckley, L., et al., 2022. Benchmarking deep learning inference of remote sensing imagery on the qualcomm snapdragon and intel movidius myriad x processors onboard the international space station. In: IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 5301–5304. http://dx.doi.org/10.1109/IGARSS46834.2022.9884906.

European Space Agency - ESA, 2018. Sentinel-2 Products Specification Document. Technical Report, pp. 48–49.

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., et al., 2017. Copernicus sentinel-2A calibration and products validation status. Remote. Sens. 9 (6), 584. http://dx.doi.org/10.3390/rs9060584.

Giglio, L., Schroeder, W., Justice, C.O., 2016. The collection 6 MODIS active fire detection algorithm and fire products. Remote Sens. Environ. 178, 31–41. http://dx.doi.org/10.1016/j.rse.2016.02.054.

Giuffrida, G., Diana, L., de Gioia, F., Benelli, G., Meoni, G., Donati, M., Fanucci, L., 2020. CloudScout: A deep neural network for on-board cloud detection on hyperspectral images. Remote. Sens. 12 (14), 2205. http://dx.doi.org/10.3390/rs12142205.

Giuffrida, G., Fanucci, L., Meoni, G., Batič, M., Buckley, L., Dunne, A., van Dijk, C., Esposito, M., Hefele, J., Vercruyssen, N., et al., 2021. The Φ-sat-1 mission: The first on-board deep neural network demonstrator for satellite earth observation. IEEE Trans. Geosci. Remote Sens. 60, 1–14. http://dx.doi.org/10.1109/TGRS.2021.3125567.

Hu, X., Ban, Y., Nascetti, A., 2021a. Sentinel-2 MSI data for active fire detection in major fire-prone biomes: A multi-criteria approach. Int. J. Appl. Earth Obs. Geoinf. 101, 102347. http://dx.doi.org/10.1016/j.jag.2021.102347.

Hu, X., Ban, Y., Nascetti, A., 2021b. Uni-temporal multispectral imagery for burned area mapping with deep learning. Remote. Sens. 13 (8), 1509. http://dx.doi.org/10.3390/rs13081509.

Kumar, S.S., Roy, D.P., 2018. Global operational land imager landsat-8 reflectance-based active fire detection algorithm. Int. J. Digit. Earth 11 (2), 154–178. http://dx.doi.org/10.1080/17538947.2017.1391341.

Lin, T., 2017. Focal loss for dense object detection. http://dx.doi.org/10.48550/arXiv.1708.02002, arXiv preprint arXiv:1708.02002.

Longepé, N., Melega, N., Marchese, V., Paskeviciute, A., Babkina, I., Petrelli, I., Casaburi, M., Peressutti, D., Kadunc, N., 2023. Φ sat-2 Mission Overview for the #ORBITALAI challenge. Technical Report, European Space Agency.

Lorenz, E., Halle, W., Fischer, C., Mettig, N., Klein, D., 2017. Recent results of the firebird mission. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 42, 105–111. http://dx.doi.org/10.5194/isprs-archives-XLII-3-W2-105-2017.

Massimetti, F., Coppola, D., Laiolo, M., Valade, S., Cigolini, C., Ripepe, M., 2020. Volcanic hot-spot detection using SENTINEL-2: A comparison with MODIS–MIROVA thermal data series. Remote. Sens. 12 (5), http://dx.doi.org/10.3390/rs12050820, URL https://www.mdpi.com/2072-4292/12/5/820.

Melega, N., Longépé, N., Marchese, V., Paskeviciute, A., Aragon, O., Babkina, I., Marin, A., Nalepa, J., Buckley, L., Guerrisi, G., et al., 2023. Implementation of the Φsat-2 on board image processing chain. In: Sensors, Systems, and Next-Generation Satellites XXVII, vol. 12729, SPIE, pp. 264–276. http://dx.doi.org/10.1117/12.2679044.

Meoni, G., Del Prete, R., Ancos-Villa, L., Albalate-Prieto, E., Rijlaarsdam, D., Espinosa-Aranda, J.L., Longépé, N., Graziano, M.D., Renga, A., 2025. E2E: Onboard satellite real-time classification of thermal hotspots events on optical raw data. Astrodynamics 1–17. http://dx.doi.org/10.1007/s42064-024-0249-x.

Meoni, G., Prete, R.D., Serva, F., De Beusscher, A., Colin, O., Longépé, N., 2024. Unlocking the use of raw multispectral earth observation imagery for onboard artificial intelligence. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 17, 12521–12537. http://dx.doi.org/10.1109/JSTARS.2024.3418891.

Mohd-Isah, N.H., Ruhaiyem, N.I.R., Abdul-Latip, N., 2020. CubeLink system-an on board data handling (OBDH) for space exploration on miniaturized satellite. In: 2020 10th IEEE International Conference on Control System, Computing and Engineering. ICCSCE, IEEE, pp. 33–36. http://dx.doi.org/10.1109/ICCSCE50387.2020.9204941.

Murphy, S.W., de Souza Filho, C.R., Wright, R., Sabatino, G., Pabon, R.C., 2016. HOTMAP: Global hot target detection at moderate spatial resolution. Remote Sens. Environ. 177, 78–88. http://dx.doi.org/10.1016/j.rse.2016.02.027.

Oktay, O., 2018. Attention U-net: Learning where to look for the pancreas. http://dx.doi.org/10.48550/arXiv.1804.03999, arXiv preprint arXiv:1804.03999.

Pu, R., Gong, P., Li, Z., Scarborough, J., 2004. A dynamic algorithm for wildfire mapping with NOAA/AVHRR data. Int. J. Wildland Fire 13 (3), 275–285. http://dx.doi.org/10.1071/WF03054.

Rijlaarsdam, D., Hendrix, T., González, P.T.T., Velasco-Mata, A., Buckley, L., Miquel, J.P., Casaled, O.A., Dunne, A., 2025. The next era for earth observation spacecraft: An overview of CogniSAT-6. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 18, 2450–2463. http://dx.doi.org/10.1109/JSTARS.2024.3509734.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp. 234–241. http://dx.doi.org/10.1007/978-3-319-24574-4_28.

Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. Superglue: Learning feature matching with graph neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4938–4947. http://dx.doi.org/10.48550/arXiv.1911.11763.

Schmitt, M., Prexl, J., Ebel, P., Liebel, L., Zhu, X.X., 2020. Weakly supervised semantic segmentation of satellite images for land cover mapping – challenges and opportunities. http://dx.doi.org/10.48550/arXiv.2002.08254, arXiv preprint arXiv:2002.08254.

Schöttl, F., Spichtinger, A., Franquinet, J., Langer, M., 2024. Real-time on-orbit fire detection on FOREST-2. In: IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 2360–2364. http://dx.doi.org/10.1109/IGARSS53475.2024.10641618.

Schroeder, W., Oliva, P., Giglio, L., Csiszar, I.A., 2014. The new VIIRS 375 m active fire detection data product: Algorithm description and initial assessment. Remote Sens. Environ. 143, 85–96. http://dx.doi.org/10.1016/j.rse.2013.12.008.

Schroeder, W., Oliva, P., Giglio, L., Quayle, B., Lorenz, E., Morelli, F., 2016. Active fire detection using landsat-8/OLI data. Remote Sens. Environ. 185, 210–220. http://dx.doi.org/10.1016/j.rse.2015.08.032.

Schulze, S.S., Fischer, E.C., Hamideh, S., Mahmoud, H., 2020. Wildfire impacts on schools and hospitals following the 2018 California camp fire. Nat. Hazards 104, 901–925. http://dx.doi.org/10.1007/s11069-020-04197-0.

Seydi, S.T., Saeidi, V., Kalantar, B., Ueda, N., Halin, A.A., 2022. Fire-net: A deep learning framework for active forest fire detection. J. Sensors 2022, 1–14. http://dx.doi.org/10.1155/2022/8044390.

Shirvani, Z., Abdi, O., Goodman, R.C., 2023. High-resolution semantic segmentation of woodland fires using residual attention unet and time series of sentinel-2. Remote. Sens. 15 (5), 1342. http://dx.doi.org/10.3390/rs15051342.

Smith, L.N., 2017. Cyclical learning rates for training neural networks. In: 2017 IEEE Winter Conference on Applications of Computer Vision. WACV, IEEE, pp. 464–472. http://dx.doi.org/10.48550/arXiv.1506.01186.

Spiller, D., Thangavel, K., Sasidharan, S.T., Amici, S., Ansalone, L., Sabatini, R., 2022. Wildfire segmentation analysis from edge computing for on-board real-time alerts using hyperspectral imagery. In: 2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering. MetroXRAINE, IEEE, pp. 725–730. http://dx.doi.org/10.1109/MetroXRAINE54828.2022.9967553.

Taha, A.A., Hanbury, A., 2015. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med. Imaging 15, 1–28. http://dx.doi.org/10.1186/s12880-015-0068-x.

Toan, N.T., Cong, P.T., Hung, N.Q.V., Jo, J., 2019. A deep learning approach for early wildfire detection from hyperspectral satellite images. In: 2019 7th International Conference on Robot Intelligence Technology and Applications. RiTA, IEEE, pp. 38–45. http://dx.doi.org/10.1109/RITAPP.2019.8932740.

Tyukavina, A., Potapov, P., Hansen, M.C., Pickens, A.H., Stehman, S.V., Turubanova, S., Parker, D., Zalles, V., Lima, A., Kommareddy, I., et al., 2022. Global trends of forest loss due to fire from 2001 to 2019. Front. Remote. Sens. 3, http://dx.doi.org/10.3389/frsen.2022.825190, URL https://www.frontiersin.org/articles/10.3389/frsen.2022.825190.

Wooster, M.J., Roberts, G.J., Giglio, L., Roy, D.P., Freeborn, P.H., Boschetti, L., Justice, C., Ichoku, C., Schroeder, W., Davies, D., et al., 2021. Satellite remote sensing of active fires: History and current status, applications and future requirements. Remote Sens. Environ. 267, 112694. http://dx.doi.org/10.1016/j.rse.2021.112694.

Wooster, M.J., Xu, W., Nightingale, T., 2012. Sentinel-3 SLSTR active fire detection and FRP product: Pre-launch algorithm development and performance evaluation using MODIS and ASTER datasets. Remote Sens. Environ. 120, 236–254. http://dx.doi.org/10.1016/j.rse.2011.09.033.

Wright, R., Flynn, L., Garbeil, H., Harris, A., Pilger, E., 2002. Automated volcanic eruption detection using MODIS. Remote Sens. Environ. 82 (1), 135–155. http://dx.doi.org/10.1016/S0034-4257(02)00030-5.

Yang, J., Zhang, Z., Wei, C., Lu, F., Guo, Q., 2017. Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4. Bull. Am. Meteorol. Soc. 98 (8), 1637–1658. http://dx.doi.org/10.1175/BAMS-D-16-0065.1.

Yu, S., Yuanbo, Y., He, X., Lu, M., Wang, P., An, X., Fang, X., 2020. On-board fast and intelligent perception of ships with the "Jilin-1" spectrum 01/02 satellites. IEEE Access 8, 48005–48014. http://dx.doi.org/10.1109/ACCESS.2020.2979476.

Zhang, Q., Ge, L., Zhang, R., Metternicht, G.I., Liu, C., Du, Z., 2021. Towards a deep-learning-based framework of sentinel-2 imagery for automated active fire detection. Remote. Sens. 13 (23), 4790. http://dx.doi.org/10.3390/rs13234790.

Zhang, B., Wu, Y., Zhao, B., Chanussot, J., Hong, D., Yao, J., Gao, L., 2022. Progress and challenges in intelligent remote sensing satellite systems. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 15, 1814–1822. http://dx.doi.org/10.1109/JSTARS.2022.3148139.