Early Wildfire Detection and Identification in High Depth-of-Field Scenarios Using LWIR Thermal Imaging from Ground-Based Systems

Antonio Galván-Hernández, Víctor Araña-Pulido, *Member, IEEE*,
Francisco Cabrera-Almeida, *Member, IEEE*, and Pedro Quintana-Morales

***Abstract*—This paper presents a novel hybrid framework that integrates spatial and temporal processing techniques for detecting and identifying incipient fires. By combining thermal imaging with a core detector, based on an object detection model, and a secondary detector, leveraging temporal features, the framework significantly enhances the detection of thermal anomalies and the identification of fires using raw Long-Wave Infrared (LWIR) thermal imaging. The framework was tested with different core detectors, trained using the Thermal Anomaly (TA) dataset, on the Fire's Latent Activity Monitoring and Evaluation through Thermography (FLAME-T) dataset, achieving improvements in mean Average Precision (mAP) and F1 scores of up to 35.9% and 20.9%, respectively, with the addition of the secondary detector. Although these improvements introduced higher processing times, the framework demonstrated its capability to maintain high detection accuracy even on a resource-constrained platform like the Raspberry Pi 5. The proposed novel identification algorithm achieves high classification accuracy for early fires at a significant depth of field, with accuracies of up to 0.913 and identification times of approximately 2 ms, making it suitable for edge applications. The code and data needed to replicate this work are available at: https://github.com/AntonioIDeTIC/IFTH.**

***Index Terms*—Incipient fire, Early detection and identification, LWIR thermal image, High depth of field.**

I. INTRODUCTION

WILDFIRES pose a global threat, requiring efficient early detection and precise monitoring systems to mitigate their impact. While the urgency of addressing this challenge is evident, effective fire detection in environments with high depth of field, such as rugged terrains and distant landscapes, remains difficult due to technological limitations and environmental influences [1].

Early wildfire detection is essential to prevent rapid spread, yet existing remote sensing approaches vary in effectiveness. Satellite-based systems, though useful for large-scale monitoring, struggle with early detection due to low temporal resolution, atmospheric interference, and detectable fire size [2], [3]. For instance, the Himawari-8 satellite provides observations every 10 minutes, while Meteosat SEVIRI updates every 5 to 15 minutes, limiting real-time detection [4], [5]. In contrast, ground-based and Unmanned Aerial Vehicle (UAV)-based systems provide faster localized detection of smaller fires before they escalate. These systems typically employ visual and thermal cameras [6], [7]. Visual sensors rely on color, motion, and texture cues but degrade under low visibility [8], while thermal cameras in the Long-Wave Infrared (LWIR) band detect fires even in smoke or low-light conditions [9]. However, several previous works on wildfire detection using ground-based or UAV-based systems have relied on 8-bit LWIR images, either processed frames or Red, Green, Blue/Infrared (RGB/IR) merges, thus losing the sensor's native radiometric resolution of 14 to 16 bits.

Recent research has advanced the field of wildfire detection by introducing diverse datasets and methodologies, each addressing specific challenges in fire monitoring and detection. Shamsoshoara et al. presented the FLAME dataset with aerial imagery of prescribed burns, achieving 92% precision and 84% recall for pixel-level fire segmentation using U-Net, and 76% accuracy in binary fire classification with a modified Xception network [10]. Chen et al. extended this with the FLAME 2 dataset, incorporating dual RGB/IR imagery to benchmark fusion methods, showing detection accuracy above 94% using state-of-the-art deep learning models [11]. Rui et al. proposed an adaptive modality learning network that improved IoU by 6.41% and F1-score by 3.39% under difficult lighting conditions [12]. Jong et al. showed that conventional datasets fail to generalize in wildfire contexts, but incorporating WIT-UAS data reduced false positives and improved accuracy with YOLOv3 and SSD, reaching an mAP of 0.566 [13].

Although many previous approaches fuse visual and thermal imagery to exploit spatial cues such as texture, color, and flicker from visible flames or smoke, they struggle in long-range, high depth-of-field settings where terrain complexity and varying distances obscure traditional fire signatures [14],

Manuscript received XXXX; revised XXXX and YYYY; accepted XXXX Date of publication XXXXX; date of current version XXXX. This work was supported by the Spanish Government under Grant PID2020-116569RB-C32 Project and under Grant PID2023-147653OB-C32 Project. Additionally, it was co-financed by the Canary Islands Agency of Research, Innovation, and Information Society of the Ministry of Universities, Science, and Innovation and Culture, and by the European Social Fund Plus (ESF+) Integrated Operational Programme for the Canary Islands 2021-2027, Axis 3 Priority Theme 74 (85%) under Grant TESIS2022010105. (Corresponding author: A. Galván-Hernández)

The authors are with the Department of Signals and Communication, Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria (ULPGC), 35017 Las Palmas, Spain (e-mail: antonio.galvan@ulpgc.es; victor.arana@ulpgc.es; francisco.cabrera@ulpgc.es; pedro.quintana@ulpgc.es).

Digital Object Identifier XXXXXX

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3644384

[15], [16]. Rugged landscapes with abrupt elevation changes, distant ravines, or mountains can mask subtle indicators, while the lack of distance information further complicates radiometric analysis. Methods tuned to short-range scenarios, such as low-flying UAVs or nearby ground cameras, benefit from fires spanning many pixels, where spatial cues remain clear [14], [17]. At larger distances, incipient fires cover only a few pixels, erasing texture and shape cues. Moreover, static heat sources such as sun-heated rocks or vehicle engines can appear as bright as distant fires, misleading networks trained on short-range data [18], [19].

Most published methods assume a zenithal viewpoint with a shallow depth of field, conditions naturally satisfied by UAVs hovering almost directly above the fire source [14], [17], [20]. Ground-based systems, however, face less favorable vantage points dictated by topography, with varying distances and frequent occlusions. The task, therefore, shifts from local fire detection to long-range fire identification, demanding new algorithms that address reduced spatial resolution and intermittent visibility. This distinction is shown in Figure 1, where position P_1 represents a close-range UAV relying on visual cues, while position P_2 shows a ground or high-altitude platform operating at much larger distances, where dedicated long-range strategies are essential.

As illustrated in Figure 1, thermal imaging can highlight hot spots via intensity thresholds, but confirming true fire activity requires radiometric analysis, which is influenced by atmospheric transmissivity, sensor calibration, and lens properties [21], [22], [23]. Different state-of-the-art studies have also demonstrated that incorporating temporal information, specifically by tracking flicker-induced volatility over time, offers a robust means of distinguishing genuine combustion dynamics from other heat sources [24], [25].

Building upon the proven efficacy of combined spatial and temporal analysis for fire detection, a significant challenge remains in reliably identifying incipient fires within complex terrains, a problem not thoroughly addressed by current state-of-the-art methods. This work proposes a hybrid framework tailored for incipient fire detection in complex terrains. The framework combines spatial and temporal analysis: a core detector based on state-of-the-art computer vision models identifies thermal anomalies with high infrared intensity peaks, while a secondary detector focuses on high temporal variability. Merging both enables reliable identification of genuine combustion events.

The framework will be executed on a platform with limited resources, such as the Raspberry Pi 5, demonstrating its feasibility on low-cost hardware. However, it is not intended for real-time operation, as the processing of multiple frames is required to capture meaningful temporal patterns. Laboratory studies indicate that early fire dynamics unfold within a relatively short observation window. Flame ignition typically stabilizes within a few seconds, followed by rapid quadratic growth of the burned area during the first two minutes, reaching sub-meter scales before slowing toward equilibrium [26]. Hotspot fires often merge with the main front within 20–40 seconds, influencing spread behavior over the following minute [27]. These findings indicate that a time window of approximately 5–120 seconds may be sufficient to capture critical fire behaviors in their early stages.

The remainder of this paper is organized as follows: Section II describes the components and methodology of the proposed fire detection and identification framework. Section III covers the datasets, implementation, hardware settings, and evaluation metrics used. Section IV presents experimental findings, and Section V concludes with key insights drawn from the results.

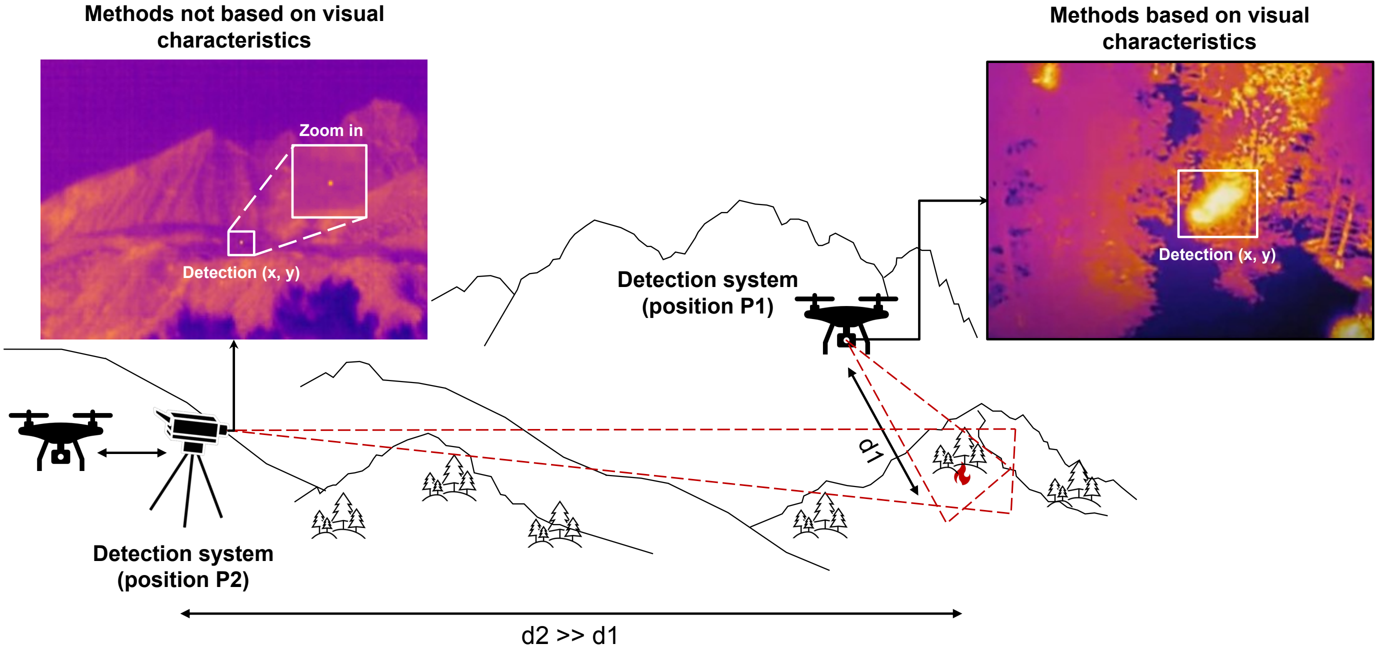


Fig. 1. Theoretical representation between two systems with different depth-of-field perspectives. Thermal image shown in position $P1$ extracted from [10].

II. METHODOLOGY

A. Overview

The proposed framework is structured into four key stages:

1) **Core Detector:** An object detection model is employed as the core mechanism for the localization of thermal anomalies, regardless of the nature of the anomaly. These anomalies are points within an image where the infrared intensity exceeds the general background values, suggesting abnormal heat sources.
2) **Secondary Detector:** The object detection model is the primary localization tool, but may miss critical early-stage fire events. The secondary detection step compensates for this by identifying regions with high temporal variability, which may correspond to early fires. These regions are also classified as thermal anomalies, as their exact cause is initially unknown.
3) **NMS Post-processing:** The results from both detectors are merged using the Non-Maximum Suppression (NMS) technique, ensuring that all critical anomalies are preserved without duplication.
4) **Identification:** The identification algorithm classifies the detected regions, determining whether the pixels correspond to potential incipient fire locations.

B. Core Detector

The Maximum Intensity Projection (MIP) metric is used on a batch of images to enhance the detection of thermal anomalies across a given time range by capturing the highest intensity values, improving potential anomaly localization. The MIP is defined as Equation 1.

$$\mathrm{MIP}(i,j) = \max_{k=1}^{N} I_k(i,j) \tag{1}$$

where $I_k(i,j)$ represents the intensity value at pixel (i,j) in the k-th frame.

In this work, models capable of object detection in a bounding box format, such as Yolov8, Real-Time Detection Transformer (RT-DETR), a hybrid Yolov8n-RT-DETR model combining Yolov8 with RT-DETR's decoder, and a custom Faster-RCNN model have been evaluated as single-class detectors for thermal imaging.

- **Yolov8:** Proposed by Ultralytics, this model uses a modified version of CSPDarknet53 as its backbone, featuring Convolution (Conv), Cross Stage Partial Bottleneck with 2 convolutions (C2f), and Spatial Pyramid Pooling Fusion (SPPF) modules. For the neck, it employs Path Aggregation Network (PANet) to enhance feature propagation, paired with C2f modules, while the head consists of Conv and C2f blocks [28], [29]. The tiny (n) version of Yolov8 is used for this implementation.
- **Yolov8-RT-DETR:** This modified Yolov8 has a similar structure but replaces the head with an RT-DETR decoder, which uses transformer-based components for more effective prediction [28].
- **RT-DETR:** Proposed by Zhao et al., it introduces HGNetV2 as its backbone, a scalable backbone that efficiently extracts multiscale features. Its neck combines the Attention-based Intra-scale Feature Interaction (AIFI) and CNN-based Cross-scale Feature Fusion (CCFF) to effectively fuse features. The head uses a transformer-based RT-DETR decoder to enhance prediction accuracy [30]. The model used has 67% less depth and has 75% reduced width compared to the original large version.
- **Faster R-CNN:** A custom Faster R-CNN model, originally proposed by Ren et al., is also tested [31]. This version adopts MobileNetV3 as the backbone, which is optimized for mobile devices through neural architecture search [32]. It uses a combination of the Region Proposal Network (RPN) and Feature Pyramid Network (FPN) as the neck and retains the original Fast R-CNN head for detection tasks [33].

These models were selected to provide a balanced and representative evaluation across different detector families and computational complexities. YOLOv8 and RT-DETR represent recently high-performing architectures widely adopted in several object detection research, offering strong baselines in both convolutional and transformer-based designs. The YOLOv8n–RT-DETR hybrid was introduced to explore the possible synergy between fast, lightweight backbones and attention-based decoding, tailored for embedded applications. Lastly, the Faster R-CNN with MobileNetV3 variant offers a well-established, low-power two-stage alternative optimized for mobile platforms.

C. Secondary Detector

This stage begins by calculating several metrics, such as the MIP, the pixel-wise variance (VAR), and the maximum Normalized Absolute Difference ($\mathrm{NAD}_{\mathrm{max}}$) for a batch of thermal images. The VAR is defined as Equation 2.

$$\mathrm{VAR} = \sigma^2(i,j) = \frac{1}{N}\sum_{k=1}^{N}\left(I_k(i,j) - \mu(i,j)\right)^2 \tag{2}$$

where $\mu(i,j)$ denotes the average intensity value at that pixel across all frames.

Meanwhile, the $\mathrm{NAD}_{\mathrm{max}}$ is computed as Equation 3.

$$\mathrm{NAD}_{\mathrm{max}}(i,j) = \max_{k=1,\ldots,N-1}\left(|I_k(i,j) - I_{k+1}(i,j)|\right) \tag{3}$$

where I_k and I_{k+1} refer to two consecutive frames. The maximum NAD is determined by identifying the highest pixel-wise difference across all NAD images.

In this work, the $\mathrm{NAD}_{\mathrm{max}}$ and VAR metrics play complementary roles in capturing different aspects of thermal variations over time. $\mathrm{NAD}_{\mathrm{max}}$ is particularly effective in detecting sudden, abrupt changes in temperature between consecutive frames, which can indicate the onset of combustion or brief thermal pulses. Conversely, VAR captures persistent thermal fluctuations across a sequence of frames, which is often characteristic of sustained combustion processes.

Secondly, as these metrics reflect the evolution of thermal intensity over time, they are combined into a composite image using a weighted sum. This image highlights regions with significant thermal changes, capturing both high peak intensity and temporal variability. Balancing these components is crucial for robust detection across diverse combustion scenarios. While some fires exhibit strong infrared emissions, making peak intensity key, others may appear weaker due to distance, noise, or obstructions. By equally weighting temporal variability and peak intensity, the composite image enhances detection reliability, as defined in Equation 4.

$$I_C = 0.5 \cdot \left(\sum_{m \in \{\mathrm{NAD_{max}}, \mathrm{VAR}\}} w_m \cdot I_m \right) + 0.5 \cdot I_{\mathrm{MIP}} \tag{4}$$

where I_C is the composite image generated. The term $\sum_{m \in \{\mathrm{NAD_{max}}, \mathrm{VAR}\}} w_m \cdot I_m$ represents the weighted sum of the temporal variability metrics. Here, w_m are the respective weights for each metric I_m, where m can be $\mathrm{NAD_{max}}$ or VAR. The term I_{MIP} is the MIP image, and the factors of 0.5 indicate that the two components of the sum are balanced equally in contributing to the final image.

Finally, the composite image is further processed to detect and localize areas of interest through the following steps:

- **Binary Thresholding:** Segments the image to isolate high-activity regions based on their intensity.
- **Region Finder:** Labels contiguous regions of pixels with similar intensities, identifying significant thermal variations. The bounding boxes, areas, and centroids are computed for each labeled region to localize these regions within thermal images.

D. NMS Post-processing

The results from the core and secondary detectors are combined using an area-based NMS algorithm, which removes overlapping bounding boxes and consolidates the final set of predictions. The process can be summarized as follows:

1) Compute the area of each bounding box and pool all detections from both detectors.
2) Sort the pooled bounding boxes in descending order of area.
3) Select the largest bounding box as the reference and compute its overlap O with each remaining box using Equation 5.

$$O = \frac{A_{\mathrm{overlap}}}{A_i + A_j - A_{\mathrm{overlap}}} \tag{5}$$

where A_{overlap} is the intersection area and A_i, A_j are the individual bounding box areas.

4) Discard any box whose overlap O exceeds the predefined threshold of 0.0001, thereby removing redundant detections. Retain all boxes with $O \leq 0.0001$ to preserve distinct anomalies.
5) Repeat steps 3–4 with the next largest remaining box until no bounding boxes remain.

This symmetric clustering and low-overlap criterion ensures that the spatial peaks of the core detector and the high-variance regions of the secondary detector are treated equally: any direct overlap is merged, while separate anomalies are kept distinct. The process directly determines the final thermal anomalies detected, each of which is interpreted as a potential fire.

E. Identification

Detecting thermal anomalies, such as spontaneous combustion, is critical when modeling long-range incipient fire scenarios. At this stage, all candidate regions have already been localized by the core and secondary detectors, and the identification phase decides whether each detected anomaly is a genuine fire or background noise. To achieve this, a novel approach is proposed to model infrared radiation changes over time across different groups of pixels. The total energy captured by the camera Y is represented as a linear combination of two components: the potential fire X_0 and the surrounding environment X_1. Here, X_1 is modeled as thermal noise, complicating the detection of X_0, as the fire occupies a smaller area and its signal diminishes with distance. To account for these factors, Equation 6 is used to model the total signal Y received by the camera as a weighted average of two thermal measurements X_0 and X_1.

$$Y = \alpha X_0 + (1 - \alpha) X_1 \tag{6}$$

where X_0 represents the potential fire, X_1 is the surrounding thermal environment. The α parameter represents the fractional contribution of X_0 to the total radiometric signal. A higher α increases detection sensitivity but may increase false alarms, while a lower α stabilizes against noise at the cost of missing weak fires.

In the absence of fire, X_0 and X_1 are assumed to be statistically independent. This assumption can be justified for long-range, high depth-of-field scenarios, since each pixel represents a large area of terrain, often hundreds of meters, so that the detected signal and its background are physically separated by considerable distances. Given these conditions, other background thermal anomalies, such as car engines, sun reflections, or natural variations, are not conductively or convectively coupled with the microenvironment of an incipient fire, and their fluctuations occur on different spatial and temporal scales that do not systematically align with the fast flicker patterns of incipient combustion at X_0.

Furthermore, during the early stages of ignition, when the fire occupies very little surface area and heat transfer is still localized, potential interactions between X_0 and X_1 may be minimal. Under this assumption, the variance of the total signal Y can be expressed as in Equation 7.

$$\mathrm{Var}(Y) = \alpha^2 \mathrm{Var}(X_0) + (1 - \alpha)^2 \mathrm{Var}(X_1) \tag{7}$$

where $\mathrm{Var}(X_0)$ and $\mathrm{Var}(X_1)$ are the variances of X_0 and X_1.

This equation defines two different stages for fire identification:

1) **Initial Phase** ($t \leq t_0$): During the initial phase, environmental noise dominates, masking the weak fire signal and making detection difficult. Under these conditions, it can be approximated that $\mathrm{Var}(Y) \approx \mathrm{Var}(X_1)$. Then, the relationship between the sources can be expressed as shown in Equation 8.

$$(2\alpha - \alpha^2) \mathrm{Var}(X_1) \approx \alpha^2 \mathrm{Var}(X_0) \tag{8}$$

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3644384

2) **Detection Phase** ($t > t_0 + t_1$): In this phase, the fire signal starts to dominate over the environmental noise, allowing confident detection. The fire identification condition can be represented as shown in Equation 9 directly from rearranging Equation 8 to isolate $\mathrm{Var}(X_0)$.

$$\mathrm{Var}(X_0) > \left(\frac{2}{\alpha} - 1\right) \mathrm{Var}(X_1) \tag{9}$$

Physically, this indicates the point at which the fire-induced variance exceeds the background variance, overcoming the slower, low-frequency drifts of non-combustible heat sources.

In this work, each detected thermal anomaly, represented by a bounding box, is evaluated individually to confirm whether it corresponds to a potential fire. Equation 9 is computed over the pixel values inside the bounding box X_0 and its surrounding background region X_1 for each frame in the LWIR image sequence. An anomaly is classified as a potential fire source only if the fire-dominant variance condition is met in at least half of the frames. This majority-frame rule ensures that transient noise is rejected, but also mitigates any minor coupling, such as slight conductive heating of background pixels by fire, by filtering out slow, correlated drifts in X_1.

The identification condition is applied using two complementary analysis strategies:

- **Global Study:** The entire image except the bounding box is treated as background noise, while the pixels inside the bounding box are considered the signal. This is appropriate when the background is uniform or the anomaly shows strong thermal contrast.
- **Local Study:** A margin is added around the bounding box, defining a local background region. The pixels inside the bounding box represent the signal, while the pixels in the surrounding margin define the background noise. This approach is particularly effective when the thermal anomaly exhibits low contrast or gradual transitions into the background.

The overall architecture of the proposed framework is shown in Figure 2. Note that the two identification conditions are applied inside the "Identification" block.

III. EXPERIMENTAL SETUP

A. Datasets

Two datasets were developed to support different aspects of this work. The first dataset, referred to in this paper as the Thermal Anomalies (TA) dataset, includes a diverse range of thermal anomalies from fires, controlled burns, urban environments, and synthetic heat sources. These scenarios simulate potential sources of false alarms, providing robust training data for detecting anomalies in real-world conditions. The images came from two main sources:

1) Previous measurement campaigns conducted by this research team, which will be made publicly available.
2) Existing state-of-the-art open datasets, such as the Advanced Driver Assistance Systems (ADAS) dataset and the M3DF dataset [34], [35].

The general characteristics of these datasets are summarized in Table I, which details the number of images, resolutions, camera models, and the variety of environments represented.

TABLE I
GENERAL INFORMATION ABOUT TA DATASET

Dataset	Images	Resolution	Camera	Environment
ADAS	242	640x512	FLIR Tau 2	Urban
M3FD	65	640x512	Not specified	Urban
Ours	969	640x480	FLIR A615	Urban, Rural & Wilderness
	548	336x256	FLIR Tau 2	
	1046	320x256	FLIR A35	
	184	320x340	Seek Mosaic	
	351	200x150	Seek Mosaic	
	1014	160x120	FLIR Lepton	

The collected TA dataset covers a wide range of fire and non-fire thermal events captured in different lighting conditions and against different backgrounds, such as urban structures, vegetation, and bare ground, to reinforce the overall robustness of the detector before specializing to fire scenarios.

Due to the absence of open datasets tracking incipient fires over time from multiple perspectives, a second dataset named Fire's Latent Activity Monitoring and Evaluation through Thermography (FLAME-T) was elaborated using a terrestrial system equipped with three LWIR thermal cameras: FLIR Tau 2, FLIR A35, and FLIR Lepton. The specifications of these cameras are summarized in Table I, while Table II provides details on the distances of the system's position relative to the fire under study. Each batch comprises 20 frames captured at 1-second intervals, for which synchronized environmental metadata was retrieved via the Open-Meteo API at timestamps as close as possible to the actual acquisition times [36]. This proof-of-concept dataset is employed to evaluate the hybrid framework developed during the measurement campaign.

TABLE II
GENERAL INFORMATION ABOUT FLAME-T DATASET

Dataset	Images*	System distance	Fire location
FLAME-T	120 x 3	Point A: 812 m	27.990619, -15.518103, 737
	140 x 3	Point B: 1130 m	
	160 x 3	Point C: 1360 m	
	120 x 3	Point D: 440 m	
	140 x 3	Point E: 1160 m	
	200 x 3	Point F: 1050 m	

* Each entry labeled "x3" indicates that images were obtained from three different cameras.

The raw LWIR thermal sensor values were captured without the camera's post-processing modifications or automatic scaling. This ensures a direct and consistent mapping between pixel intensity and infrared radiation, preserving the integrity of the measurements across frames and removing variations introduced by dynamic camera adjustments.

Figure 3 shows a fire of the FLAME-T dataset captured with the three cameras discussed above, highlighting the impact of resolution (R), Horizontal Field of View (HFOV), and Vertical

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3644384

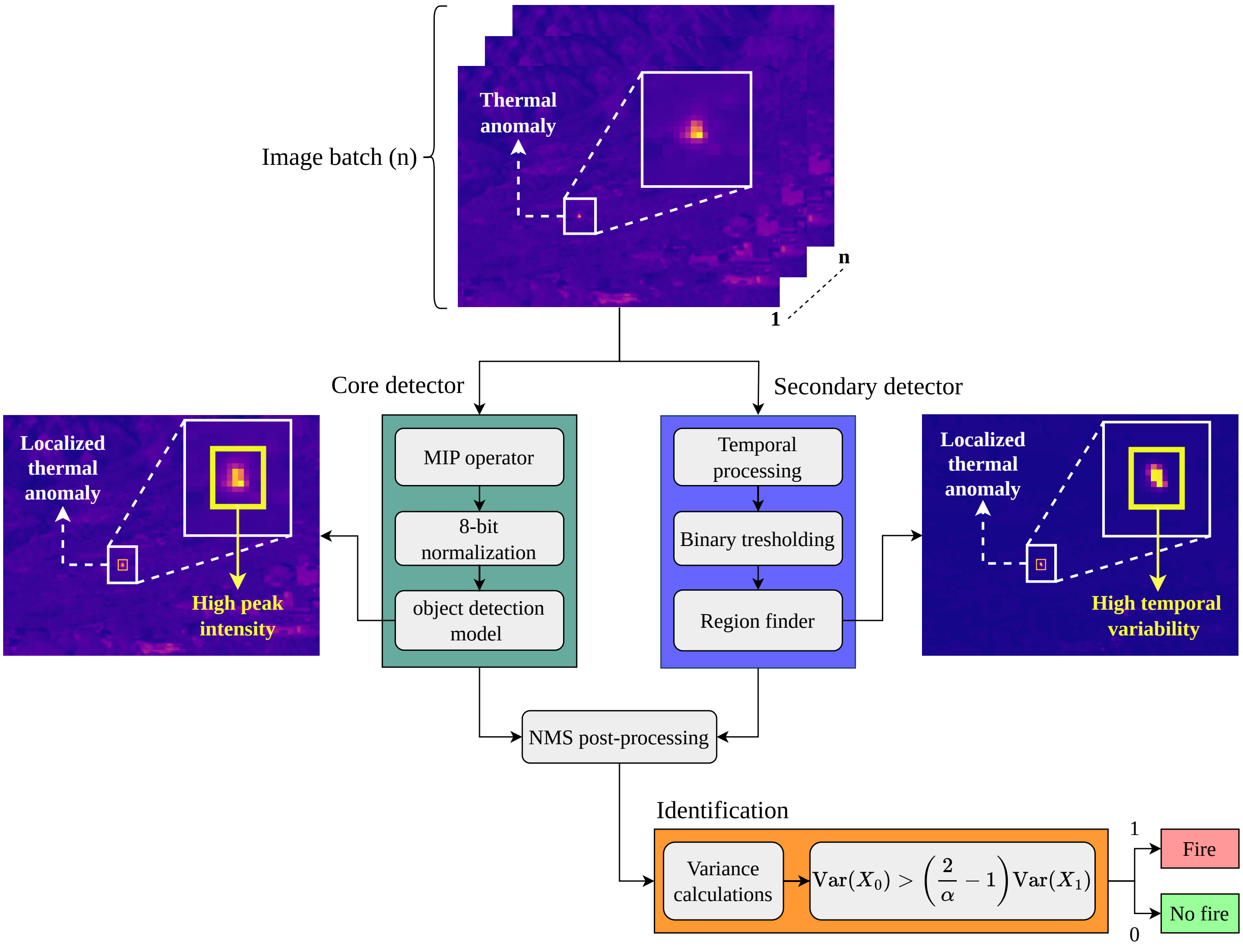


Fig. 2. Proposed framework for early fire detection and identification. Core detector (green), Secondary detector (blue), and Identification (orange).

Field of View (VFOV) on image representation. In addition, a zoomed-in view of the heat source, enclosed in a white box, is provided. Thermal measurements were collected from each camera and overlaid to demonstrate the variability in the data:

- **Coefficient of Variation (CV):** Provides a standardized measure of dispersion relative to the mean. It is calculated as the ratio of the standard deviation to the mean, as shown in Equation 10.

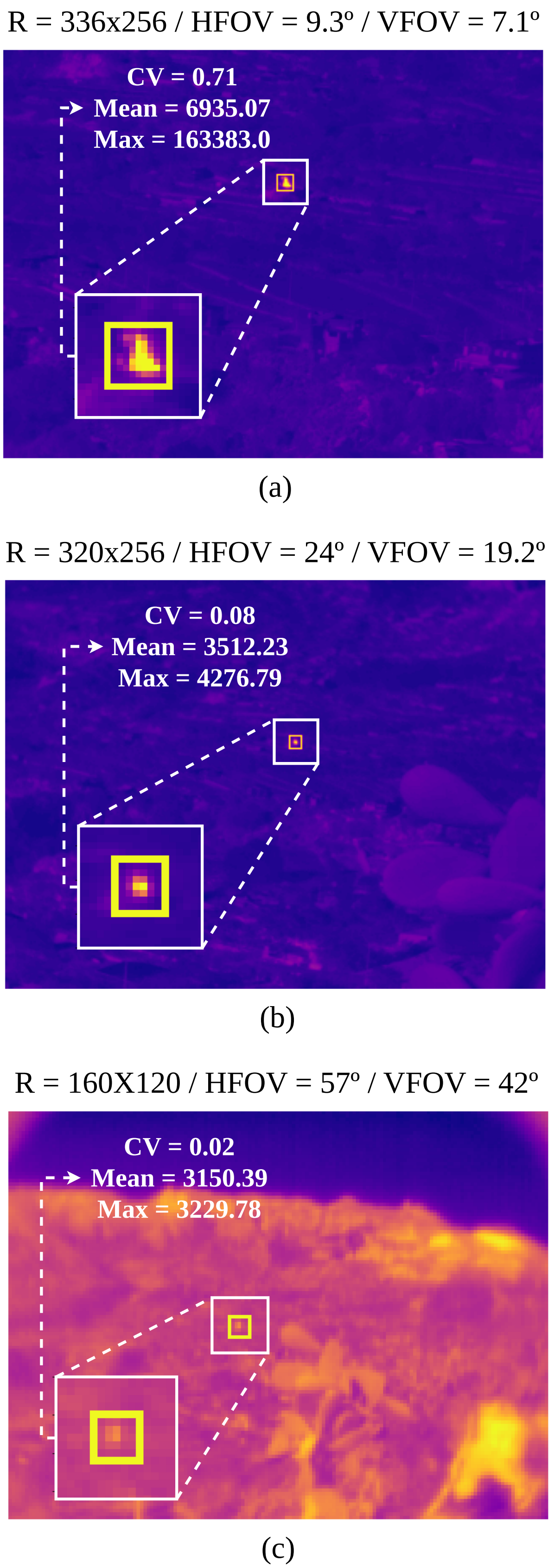
$$CV(i,j) = \frac{\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(I_k(i,j) - \mu(i,j)\right)^2}}{\frac{1}{N}\sum_{k=1}^{N} I_k(i,j)} \tag{10}$$

- **Mean and Max values:** Represent the average and peak raw thermal counts captured by the cameras.

In Figure 3(a), the thermal image was captured with a Tau 2 camera, featuring a resolution of 336x256 and a very narrow HFOV (9.3°) and VFOV (7.1°). This configuration provides precise but highly variable readings, as reflected in a high CV (0.71) and widely spread intensity values (mean: 6935.07, max: 163383.0). Figure 3(b) shows an image captured with a A35 camera, which has a similar resolution (336x256) but a wider HFOV (24°) and VFOV (19.2°). The readings are more stable in this case, with a low CV (0.08) and balanced intensity values (mean: 3512.23, max: 4276.79). In contrast, Figure 3(c) presents an image captured with a Lepton camera, offering a lower resolution of 160x120 and the widest HFOV (57°) and VFOV (42°). This setup prioritizes broader coverage over detail, resulting in very stable readings characterized by a minimal CV (0.02) and closely grouped intensity values (mean: 3150.39, max: 3229.78).

These measurements highlight how camera specifications and calibration influence the detection and interpretation of fire characteristics. While proper distance calibration could enable a general fire identification algorithm to perform consistently across different cameras, uncalibrated setups require camera-specific adjustments such as tailored variables or thresholds to optimize detection accuracy, especially given that the Analog-to-Digital Converters (ADCs) in each camera digitize the intensity of infrared radiation in different ways.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3644384



(a)

(b)

(c)

Fig. 3. The same burning area captured by different LWIR thermal cameras. (a) FLIR Tau 2 camera. (b) FLIR A35 camera. (c) FLIR Lepton camera.

B. Implementation Details

The TA dataset, shown in Table I, was split 70%-30% for training and evaluation, resulting in 3102 and 1330 images, respectively. To enhance generalization, pixel-level (e.g., random blur, noise, contrast enhancement) and spatial-level (e.g., rotations, shifts, flips) transformations were applied to the training data, increasing from 3102 to 9279 images.

The optimization of the proposed framework was conducted through a two-stage grid search procedure:

- The first stage focused on configuring the parameters for generating the composite image defined in Equation 4. This process involved selecting suitable weights for each component ($w_{\text{NAD}}, w_{\text{VAR}}$), along with appropriate binarization thresholds. The resulting optimal values were $w_{\text{NAD}} = 0.55$ and $w_{\text{VAR}} = 0.25$. For effective isolation of significant features, the final binary mask was thresholded at 40% of the maximum pixel value for Tau 2 images and at 10% for both A35 and Lepton images.
- The second stage involved a separate grid search to determine the optimal α values and margin sizes for the identification process described in Equation 9. The α parameter was optimized independently for both the global study and the local study. In contrast, the margin was only relevant and optimized for the local study, where a region around the detected anomaly is analyzed.

Finally, the core and secondary detections are merged using the NMS post-processing step, and the identification process is applied. The normalized variance, calculated as the variance of pixel values divided by their mean, is employed to analyze and interpret the results presented.

C. Tests on Different Hardware

Modern computer vision models can be executed remotely on external servers, but in wildfire scenarios with steep and rugged terrain, communication failures and bandwidth limitations are common [37]. To address these challenges, a resource-constrained hardware platform was evaluated to identify a methodology suitable for edge processing. Edge processing enables the transmission of processed results instead of raw images, conserving bandwidth and mitigating communication issues [38]. Crucially, by performing all detection and identification locally on the edge device, the system remains operational even when connectivity is poor or intermittent, ensuring no loss of critical early-warning capability.

The computational efficiency of the framework was evaluated on two different platforms: a high-performance training workstation and a Raspberry Pi 5 (4GB). The object detection models were trained in the workstation with an NVIDIA 2080Ti Graphics Processing Unit (GPU), Intel i9-9900X @ 3.50GHz, and 32 GB of RAM. For the Raspberry Pi, the models were optimized using Open Neural Network Exchange (ONNX) to boost inference speed and efficiency for framework evaluation purposes.

D. Evaluation Metrics

The detection performance is evaluated using standard machine-learning metrics like the mAP at IoU thresholds of

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3644384

0.5 and across the 0.5–0.95 range ($\mathbf{mAP_{0.5}}$, $\mathbf{mAP_{0.5:0.95}}$), alongside the F1 score, while the model's ability to distinguish fire from non-fire is quantified by classification accuracy. However, as detection of thermal anomalies and identification of potential fires use different criteria for true positives (TP), false positives (FP), and false negatives (FN), these conditions are summarized in Equation 11 and Equation 12.

$$\text{Detection} = \begin{cases} \mathbf{TP}, & \text{if IoU} \geq \text{Th}, \\ \mathbf{FP}, & \text{if IoU} < \text{Th}, \\ \mathbf{FN}, & \text{if missed detection}, \end{cases} \tag{11}$$

$$\text{Identification} = \begin{cases} \mathbf{TP}, & \text{if IoU} \geq \text{Th} \wedge \text{Condition}, \\ \mathbf{FP}, & \text{if IoU} < \text{Th} \wedge \text{Condition}, \\ \mathbf{FN}, & \text{otherwise}. \end{cases} \tag{12}$$

where $\wedge$ is the AND operator, Th is the chosen threshold, and the condition is described in Equation 9.

IV. RESULTS AND ANALYSIS

This section evaluates the proposed framework through different experiments, each designed to assess different aspects of its performance:

1) **Benchmarking Object Detection Models:** The first experiment involves training and benchmarking different object detection models as single-class detectors using the validation set described in Table I. These models are trained exclusively on the TA dataset, which includes a diverse range of thermal anomaly scenarios beyond wildfires. The goal of this experiment is to establish a baseline performance reference for the models, ensuring reproducibility and facilitating comparisons with future research. The trained models from this stage are used in subsequent evaluations.
2) **Secondary Detector Ablation Experiment:** The ablation experiments were performed using the FLAME-T dataset by varying the frame window size in the secondary detector and testing different metric configurations to generate the composite image. These studies help to evaluate the trade-offs between inference time and detection accuracy in different configurations.
3) **Framework Detection Analysis:** In this experiment, the framework is also evaluated on the FLAME-T dataset. This approach assesses how well the models, originally trained on a broader thermal anomaly dataset, adapt to fire-related scenarios. Moreover, the secondary detector is expected to reduce missed detections, thereby improving the mAP and F1 scores.
4) **Framework Identification Analysis:** This experiment extends the previous one by combining the results from both the core and secondary detectors as input for the fire identification process. This step aims to refine fire classification further by leveraging the strengths of both detection mechanisms, ensuring improved robustness and reliability in wildfire detection.
5) **Comparison With State-of-the-Art Methods:** The proposed framework is compared with other state-of-the-art approaches, considering key factors such as image type, studied scenario, data source, hardware implementation, deployment on embedded systems, and associated costs.

A. Benchmarking Object Detection Models

The results are summarized in Table III, which includes metrics for object detection models trained on the TA dataset. Inference times, presented in milliseconds (ms), are distinguished by platform: red for the training platform and blue for the Raspberry Pi 5. The best-performing values for each metric are highlighted in bold. In addition to the $\mathbf{mAP_{0.5}}$ and F1 score, the $\mathbf{mAP_{0.5:0.95}}$ is also reported to provide a more comprehensive evaluation of model performance across different IoU thresholds. Furthermore, Giga Floating Point Operations Per Second (GFLOPs) and the number of parameters are included to assess complexity and computational efficiency.

For the TA validation images, it was observed that converting and optimizing the models to ONNX had no significant impact on the overall performance. Among the evaluated models, YOLOv8n demonstrated the best results, achieving the highest $\mathbf{mAP_{0.5}}$ (**0.692**) and F1 score (**0.678**), along with a higher $\mathbf{mAP_{0.5:0.95}}$ (**0.320**), while maintaining relatively low GFLOPs and parameters. This suggests that YOLOv8n offers an effective balance between accuracy and computational efficiency, making it well-suited for thermal anomaly detection on resource-constrained platforms, and possesses strong generalization capabilities for detecting diverse types of thermal anomalies.

B. Secondary Detector Ablation Experiment

The results are summarized in Table IV. The term 'Detection time' is used since the secondary detector is not based on a neural network model. Note that the results are highlighted as in the previous experiment. The frame window refers to the number of images used.

These findings indicate that increasing the frame window size does not necessarily improve detection accuracy when NAD_{max} and VAR are used independently. Performance may plateau or slightly decline beyond a certain number of frames. However, when both metrics are combined using the weight configurations discussed in subsection III-B, the integration of their complementary information leads to improved detection performance, obtaining the highest accuracy observed in a window size of 20 frames, with a $\mathbf{mAP_{0.5}}$ of **0.700** and an F1 score of **0.663**. It is important to emphasize that this improvement is based on empirical results with the current dataset and cannot be generalized to larger frame windows without additional data. The behavior of fire beyond this observation window may vary significantly due to factors such as wind, humidity, fuel conditions, and terrain. As such, the transition from early fire detection to tracking fire spread requires further investigation.

In this work, the 20-frame window was chosen for the remainder of the experiments as a practical tradeoff between temporal coverage and processing limitations, especially for resource-constrained platforms such as the Raspberry Pi 5,

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3644384

TABLE III
OBJECT DETECTION PERFORMANCE FOR TA DATASET

Hardware settings
Training platform / Raspberry Pi 5

Model	**Inference time**	**mAP$_{0.5}$**	**mAP$_{0.5:0.95}$**	**F1 score**	**GFLOPs**	**Parameters (M)**
Yolov8n	6 / 162	**0.692**	**0.320**	**0.678**	7.1	≈ 2.724
Yolov8n-RT-DETR	21 / 461	0.527	0.190	0.623	17.1	≈ 9.643
RT-DETRn	27 / 818	0.586	0.209	0.662	38.2	≈ 15.680
Faster R-CNN	11 / 379	0.496	0.200	0.496	4.49	≈ 19.386

TABLE IV
ABLATION EXPERIMENT FOR THE SECONDARY DETECTOR

Hardware settings
Training platform / Raspberry Pi 5

	NAD$_{max}$				**VAR**			
Frame window	**Detection time**	**mAP$_{0.5}$**	**mAP$_{0.5:95}$**	**F1 score**	**Detection time**	**mAP$_{0.5}$**	**mAP$_{0.5:95}$**	**F1 score**
4	13 / 19	**0.391**	0.105	**0.298**	10 / 35	0.434	0.138	0.404
8	17 / 49	0.343	0.117	0.260	18 / 56	**0.621**	0.241	0.597
12	22 / 67	0.359	0.115	0.272	21 / 73	0.589	**0.253**	0.570
16	27 / 88	0.374	**0.121**	0.284	26 / 86	0.581	0.228	0.565
20	33 / 108	0.343	0.120	0.260	37 / 126	0.615	0.252	**0.599**

	NAD$_{max}$ & VAR			
Frame window	**Detection time**	**mAP$_{0.5}$**	**mAP$_{0.5:95}$**	**F1 score**
4	12 / 47	0.314	0.093	0.278
8	23 / 78	0.606	0.257	0.567
12	39 / 129	0.643	0.272	0.608
16	47 / 156	0.643	0.271	0.608
20	57 / 178	**0.700**	**0.292**	**0.663**

where computational overhead is increasingly significant. Although this configuration provides the most robust detection observed in the experiments, its feasibility in real-world edge applications must still account for latency requirements and hardware limitations.

C. Framework Detection Analysis

The results of the thermal anomalies detection stage are summarized in Table V. The best-performing model is highlighted based on the platform: red for the training platform and blue for the Raspberry Pi 5. The table is divided into two parts: one for results from the core detector alone and the other for results combining the core and secondary detectors using the NMS algorithm.

The object detection models were initially trained and evaluated on a dataset containing various thermal anomaly scenarios. In contrast, the framework evaluation conducted with the FLAME-T dataset focused specifically on fire scenarios. This difference in data distribution likely explains some of the observed performance discrepancies, particularly for ONNX-converted models. Furthermore, the evaluation approaches differed: in the first experiment, models were tested on individual images, while in the second experiment, batches of images were processed to generate composite images using MIP and temporal metrics. This shift in evaluation methodology likely influenced the performance results, as batch processing introduces additional factors, such as aggregated temporal features, that may affect the performance results.

The evaluation of models on the FLAME-T dataset underscores two key findings: the substantial accuracy improvements achieved by integrating a secondary detector and the trade-offs between accuracy and processing time across hardware platforms. RT-DETRn emerged as the most accurate model, achieving the highest mAP$_{0.5}$ with **0.788** and notable F1 scores, such as **0.821-0.819**, using only the core detector. When combined with the secondary detector, its mAP$_{0.5}$ improved to **0.799-0.800**, and F1 score surged to **0.866-0.868**, reflecting its ability to focus on specific patterns associated with fire detection, especially in datasets with simpler decision boundaries, such as those containing only fires. Similarly, YOLOv8n exhibited significant gains, with its mAP$_{0.5}$ increasing from **0.695-0.723** to **0.750-0.788** and its F1 score from **0.781-0.793** to **0.826-0.840**, demonstrating that the secondary detector effectively enhances the detection of thermal anomalies. It should be noted that the Faster R-CNN model experienced the most substantial improvement, with mAP$_{0.5}$ rising from **0.541-0.545** to **0.735-0.744** and F1 scores increasing from **0.604-0.619** to **0.730-0.737**, highlighting the

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3644384

TABLE V
FRAMEWORK ANALYSIS FOR FLAME-T DATASET (I)

Hardware settings								
Training platform / Raspberry Pi 5								
	Core detector				**Core detector + secondary detector**			
Model	**It**[1]	**mAP$_{0.5}$**	**mAP$_{0.5:0.95}$**	**F1 score**	**Tdt**[1,2]	**mAP$_{0.5}$**	**mAP$_{0.5:0.95}$**	**F1 score**
Yolov8n	7 / 187	0.695 / 0.723	0.289 / 0.298	0.781 / 0.793	57 / 360	0.750 / 0.788	0.308 / 0.347	0.826 / 0.840
Yolov8n-RT-DETR	22 / 524	0.761 / 0.749	0.262 / 0.264	0.822 / 0.813	73 / 698	0.773 / 0.769	0.290 / 0.292	0.837 / 0.835
RT-DETRn	29 / 791	0.788 / 0.788	0.308 / 0.306	0.821 / 0.819	79 / 964	0.799 / 0.800	0.308 / 0.308	0.866 / 0.868
Faster R-CNN	13 / 370	0.541 / 0.545	0.185 / 0.191	0.604 / 0.619	63 / 544	0.735 / 0.744	0.302 / 0.304	0.730 / 0.737

[1] It stands for Inference time, and Tdt stands for Total detection time.
[2] Tdt includes inference time from the core detector and detection time from the secondary detector using a frame window of 20 images.

effect of the secondary detector on models with initially lower detection performance.

Furthermore, the improvements are mostly consistent with the more stringent metric mAP$_{0.5:0.95}$, highlighting the secondary detector's ability to recover missed detections at different IoU levels. While RT-DETRn's mAP$_{0.5:0.95}$ score remains similar, its F1 score experiences a substantial jump, highlighting a significant reduction in missed detections. Similarly, mAP$_{0.5:0.95}$ of YOLOv8n increases from **0.289-0.298** to **0.308-0.347**, which seems to indicate that the study of the variability between frames improves the localization of subpixel fires that the spatial detector alone would not have been able to detect.

However, the addition of the secondary detector also introduced significant processing time increases, particularly on resource-limited platforms like the Raspberry Pi 5. For instance, YOLOv8n's detection time increased from **187 ms** to **360 ms**, while RT-DETRn's detection time grew from **791 ms** to **964 ms**. Similarly, Faster R-CNN experienced a processing time increase from **370 ms** to **544 ms**, highlighting the computational costs of improved accuracy.

The total detection time of the framework using YOLOv8n, including both core and secondary detectors, is approximately half that of RT-DETRn alone while achieving comparable mAP and an improved F1 score due to a reduction in false negatives by implementing the secondary detector. This makes lightweight models like YOLOv8n, when paired with an additional detector, a more efficient and effective solution, particularly for edge applications on embedded systems like the Raspberry Pi 5, where computational efficiency is critical.

D. Framework Identification Analysis

The results of the framework identification stage are summarized in Table VI, which presents identification time, alpha values, and accuracy for global and local studies. These results highlight the performance improvements obtained by optimizing parameters for each camera. In this work, the YOLOv8n model was chosen as the baseline core detector to illustrate the proposed pipeline end-to-end, as it provides a balanced trade-off between detection time, mAP, and F1 score. Beyond its lighter architecture and favorable compute/memory footprint, YOLOv8n also benefits from mature embedded-system tooling, making it more practical for integration on platforms such as the Raspberry Pi 5. Importantly, YOLOv8n achieves competitive accuracy, including similar or even better performance than RT-DETRn in more general metrics such as mAP$_{0.5:0.95}$.

The local study improved identification accuracy for thermal images captured by the Lepton and A35 cameras. For the Lepton images, where fires are represented by very few pixels with low variance, accuracy increased from **0.162** to **0.578** on the training platform and from **0.162** to **0.595** on the Raspberry Pi 5. Similarly, for the A35 images, accuracy improved from **0.822** to **0.913** across both platforms, demonstrating the advantages of localized pixel analysis. Conversely, for the Tau 2, accuracy remained stable at **0.87** across both global and local studies. While the Tau 2 shares the same resolution as the A35, its narrower FOV enhances thermal change detection by focusing on a smaller region. This increases pixel-level detail and temporal variations, making it easier for the identification algorithm to detect fire-related anomalies with high precision.

In addition to accuracy improvements, the local study significantly reduced identification times for processing thermal images from all cameras. For the Lepton, identification time decreased from **26 ms** to **1 ms** on the training platform and from **73 ms** to **2 ms** on the Raspberry Pi 5. For the A35, identification times dropped from **26 ms** to **2 ms** on the training platform and from **73 ms** to **3 ms** on the Raspberry Pi 5. Even for the Tau 2, where accuracy remained unchanged, identification time reduced from **26 ms** to **7 ms** on the training platform and from **73 ms** to **11 ms** on the Raspberry Pi 5. These findings emphasize that the local study improves both accuracy and processing efficiency, particularly for thermal images captured by cameras like the Lepton and A35, making it a practical approach for resource-constrained environments.

The condition for correct early fire identification can be modified depending on the scenario, as it is based on Equation 12. For illustrative purposes, Figure 4 shows different accuracy results over different core detector confidences. In contrast to Table VI, the accuracy values are represented as the average of the results obtained from all cameras. The core detector exhibits minimal accuracy for confidence levels above **0.8** while the secondary detector remains operational. This scenario simulates real-world conditions where the core detector fails to detect anomalies. Yet, the secondary detector successfully detects them, enabling the identification algorithm to discern whether it is an incipient fire or not.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3644384

TABLE VI
FRAMEWORK ANALYSIS FOR FLAME-T DATASET (II). OPTIMIZED α AND, WHERE APPLICABLE, MARGIN VALUES WERE DETERMINED VIA GRID SEARCH FOR EACH CAMERA AND STUDY TYPE TO ENHANCE IDENTIFICATION ACCURACY.

Hardware settings **Training platform / Raspberry Pi 5**								
		Global study			**Local study**			
Camera	**Total fires**	**Identification time[1]**	α	**Accuracy$_{0.5}$**	**Identification time**	α	**Margin[2]**	**Accuracy$_{0.5}$**
FLIR Tau 2	43	26 / 73	0.2	0.87 / 0.87	7 / 11	0.15	155	0.87 / 0.87
FLIR A35	44	26 / 73	0.3	0.822 / 0.822	2 / 3	0.25	25	0.913 / 0.913
FLIR Lepton	37	26 / 73	0.6	0.162 / 0.162	1 / 2	0.45	10	0.578 / 0.595

[1] Same global identification times as all images were resized to 640x640.
[2] Number of pixels added around the thermal anomaly's bounding box that defines the thermal noise.

Finally, after presenting and analyzing all the conducted experiments, the complete workflow of the framework is illustrated in Figure 5, providing a comprehensive overview of the detection and identification process. This visualization highlights the progression of thermal anomaly detection and early fire identification in a high depth-of-field scenario.

E. Comparison With State-of-the-Art Methods

To the best of the authors' knowledge, no publicly available method or dataset is specifically designed for detecting early incipient fires at long distances using only LWIR images in high depth-of-field scenarios. As a result, direct numerical comparisons with existing methods are not strictly valid. Most existing approaches, such as those listed in Table VII and Table VIII, rely on visible fire and smoke features. In contrast, the proposed method addresses:

1) **High Depth-of-Field Scenarios:** Existing methods assume a relatively shallow depth of field and short distances, exploiting visible cues of fire or smoke. This work targets long-range detection where these cues are often imperceptible in visible and infrared imagery.
2) **LWIR Imagery:** Since early fire and smoke features are not visible in the optical spectrum at large distances, this work relies solely on raw infrared data. This introduces additional challenges regarding resolution and scene complexity. Unlike existing approaches that rely on 8-bit or post-processed thermal frames, the proposed framework operates directly on raw 16-bit LWIR information, preserving full radiometric resolution for the detection of very small, distant fires.
3) **Lack of Existing Datasets:** No current public dataset exhibits these high depth-of-field, early fire characteristics in LWIR; hence, this work proposes a new dataset (FLAME-T) specifically collected for such cases. This dataset is a novel contribution to the community, enabling future research in long-range early fire detection.

Due to these methodological differences, the comparison in Table VII and Table VIII focuses on broader attributes, including data type and resolution used, dataset used, hardware resources, runtime performance, tested embedded deployment, and overall cost.

The proposed framework differentiates itself from existing state-of-the-art wildfire detection methods by prioritizing lightweight, cost-effective, and deployable solutions, addressing key limitations in prior works. Most existing approaches rely on RGB-based fire detection with high-resolution images [17], [40], [44] or processed IR data that enhance thermal contrasts but lose vital radiometric information [14], [41], [42]. By working directly with raw LWIR images, this framework preserves crucial thermal details, enabling the detection of early fires, even in high depth-of-field scenarios where fire and smoke features are not easily discernible.

Another key distinction lies in hardware requirements and deployment feasibility. Many state-of-the-art models are trained and deployed on high-performance GPUs, such as Tesla A100, Tesla V100, and NVIDIA A800 [14], [40], [44], or consumer-grade GPUs like the RTX 3070, RTX 2080 Ti, and 1080 Ti [17], [42], [43]. While these models achieve impressive accuracy, they are often impractical for real-world deployment in remote or resource-constrained environments. Some efforts mitigate this by training on high-end GPUs while deploying on embedded platforms like Jetson Xavier NX or AGX Xavier [20], [41], but these still present cost and power challenges, limiting scalability. In contrast, the proposed framework is designed for affordable, low-power deployment using a Raspberry Pi 5, demonstrating its feasibility on a cost-effective and energy-efficient platform, as it completes the entire processing pipeline in less than 25 seconds, within the estimated early fire observation window discussed in different state-of-the-art studies [26], [27]. As a result, it becomes possible to transmit only compact detection results, such as bounding box coordinates, confidence scores, and location information, thus minimizing bandwidth usage.

V. CONCLUSION

This work presents a novel framework for detecting and identifying early-stage fires in complex environments with significant depth of field using LWIR thermal imaging. By integrating a dual-detection mechanism that captures both spatial and temporal variations alongside a novel fire identification strategy, the proposed framework achieves improved accuracy and reliability in fire detection and identification.

Quantitative results validate the benefits of this approach. On the proposed FLAME-T dataset, thermal anomaly detection performance was evaluated using both the core detector alone and the combined core detector with the secondary

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3644384

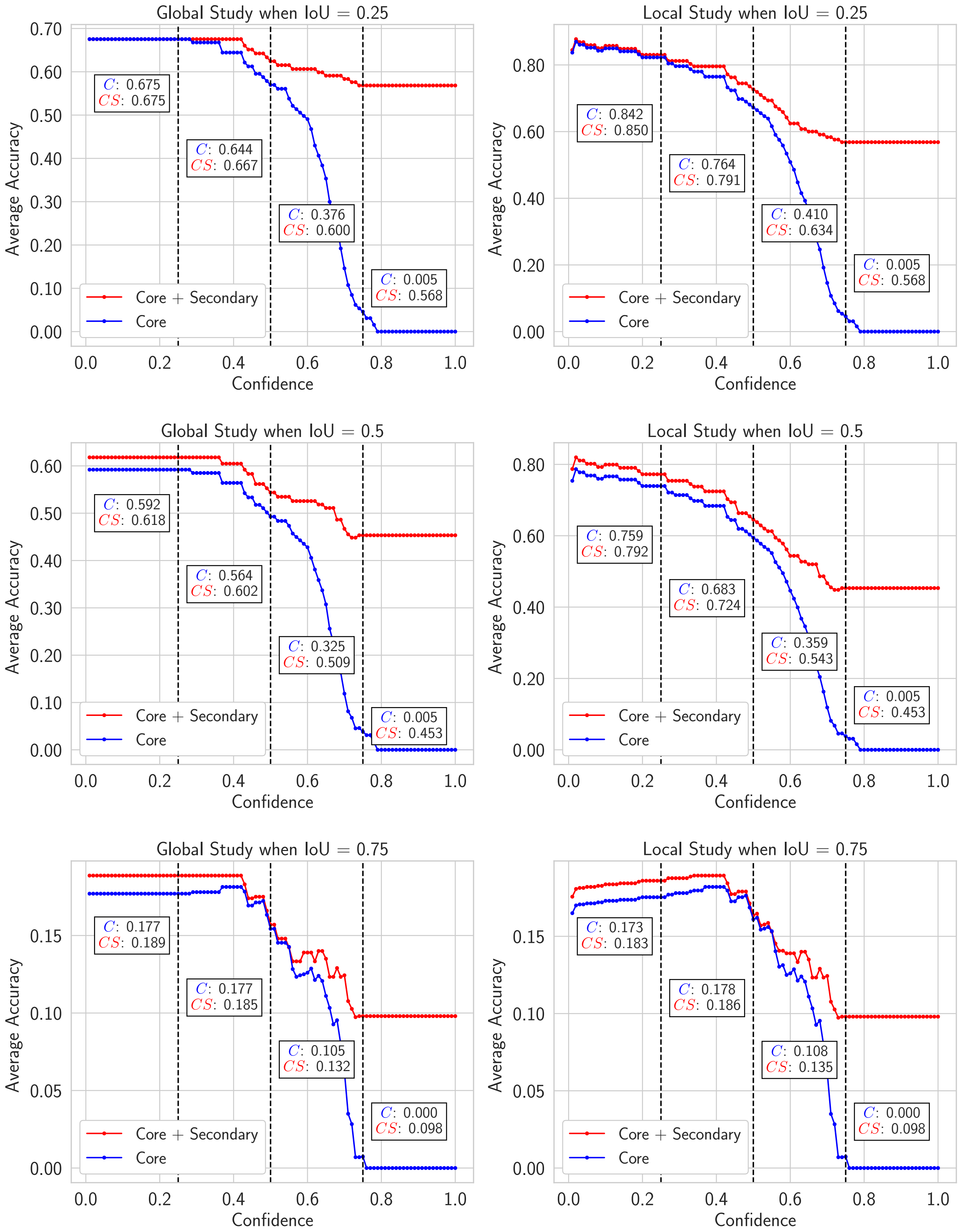


Fig. 4. Accuracy results for different IoU. First, second, and third rows with IoU greater than 0.25, 0.5, and 0.75, respectively. First column for the global study and second column for the local study. The boxes between the dashed vertical lines provide average accuracy data for each confidence interval.

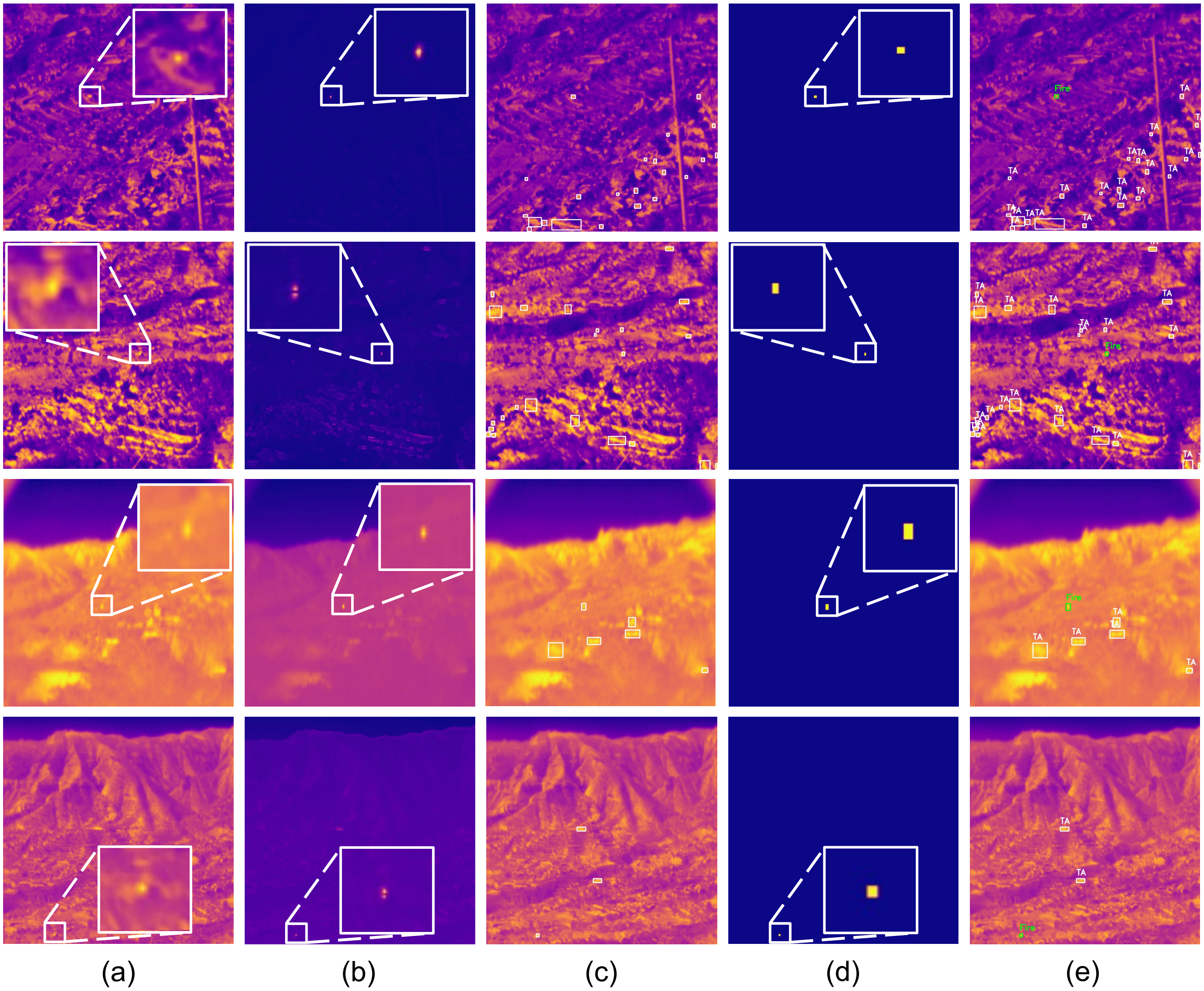


Fig. 5. Framework workflow for early fire detection and identification. (a) Original image (first one in the batch). (b) Composite image before binarization. (c) Thermal anomalies detection after NMS post-processing. (d) Fire identification algorithm applied to a bounding box. (e) Final visual result.

detector. Integrating the secondary detector significantly improved the overall framework performance. For instance, RT-DETRn showed a 1.4% mAP$_{0.5}$ increase and a 5.5% improvement in F1 scores when the secondary detector was added. Similarly, YOLOv8n achieved a 7.9% mAP$_{0.5}$ increase and a 5.8% improvement in F1 scores by incorporating the secondary detector. Notably, Faster R-CNN exhibited the highest relative improvement, with a 35.9% increase in mAP$_{0.5}$ and a 20.9% boost in F1 scores, demonstrating how the secondary detector effectively mitigates the model's initial limitations. While the framework demonstrated significant improvements in detection performance, these gains came with increased processing times on the Raspberry Pi 5. This highlights the trade-off between accuracy and efficiency in deploying the framework for edge applications.

The inclusion of both global and local studies provided valuable insights into the adaptability of the identification algorithm across various thermal cameras. The local study, in particular, proved effective in optimizing classification accuracy and efficiency, especially for thermal cameras like the Lepton and A35, making this identification algorithm practical for edge deployments in diverse and challenging environments.

The complete processing of 20 frames, including image acquisition, detection, and identification, takes less than 25 seconds on the Raspberry Pi 5, falling within the estimated early-fire observation window of approximately 5–120 seconds, allowing key detection information to be transmitted while reducing bandwidth usage, as the processing is performed on-edge. It should be noted that in applications covering large areas, such as scenarios requiring multiple perspectives or the use of wide-angle cameras, processing times may vary due to an increased number of pixels per frame and expanded FOV. In such cases, computational requirements may rise, particularly when handling higher-resolution thermal data. Nevertheless,

TABLE VII
COMPARATIVE TABLE WITH OTHER STATE-OF-THE-ART APPROACHES FOR WILDFIRE DETECTION (I): RESOLUTION (R), SOURCE (S)

Method	**Data Type**	**Dataset**	**Scenery Type**
GMM-EM [39]	Video (RGB) R: 320×240 to 1600×1200	VisiFire, FireSense S: Wildfires, Smoke, Candles, etc.	Ground Short Distances
Miti-DETR [40]	Images (RGB) R: Unknown	FLAME S: Controlled Fire, Smoke	UAV Zenithal Perspective Shallow Depth of Field
Inv-Net [17]	Images (RGB) R: 3840×2160		
AMSO-SFS [41]	Images (RGB + Processed LWIR + SAR) R: 512×512	FLAME, FLAME2 S: Controlled Fire, Smoke	
CT-Fire & DC-Fire [42]	Images (RGB + Processed LWIR) R: 224×224		
Encoder-Decoder [14]	Video (RGB + Processed LWIR) R: Unknown	FLAME2 S: Controlled Fire, Smoke	
FFS-Unet [43]	Images (RGB) R: 448×448	FLAME, ERA, Corsican Fire S: Controlled Fire, Wildfire, Smoke	Ground, UAV Zenithal Perspective Shallow Depth of Field
RFWNet [20]	Images (RGB) R: 448×448 and 832×832	DFS-FIRE, Proprietary, Synthetic S: Wildfire, Smoke, etc.	Ground, UAV Mixed Perspectives
Custom YOLOv8 [44]	Images (RGB) R: Unknown	D-Fire, Proprietary S: Wildfire, Large Smoke Plumes	UAV Mixed Perspectives
Ours	Images (Raw LWIR) R: 640×640	FLAME-T S: Very small burning area	Ground High Depth of Field Different Positions

TABLE VIII
COMPARATIVE TABLE WITH OTHER STATE-OF-THE-ART APPROACHES FOR WILDFIRE DETECTION (II)

Method	**Hardware**	**Runtime**[1]	**Deployment** [2]	**Cost**[3]
GMM-EM [39]	Intel Core i3 (inference)	≈ 48 ms	Yes	Low
Miti-DETR [40]	Tesla V100 (train + inference)	≈ 9.8 ms	No	High
Inv-Net [17]	GTX 3070 (train + inference)	≈ 640 ms	No	High
AMSO-SFS [41]	RTX 3090 (train) Jetson AGX Xavier (inference)	≈ 54 ms (Jetson)	Yes	Medium
CT-Fire & DC-Fire [42]	RTX 2080 Ti (train + inference)	DC-Fire: ≈ 13 ms CT-Fire: ≈ 24 ms	No	High
Encoder-Decoder [14]	Tesla A100 (train + inference)	Unknown	No	High
FFS-Unet [43]	GTX 1080Ti (train + inference)	≈ 20 ms	No	High
RFWNet [20]	NVIDIA RTX 4080 (train + inference) Jetson Xavier NX (inference)	≈ 3.26 ms (Desktop) ≈ 50 ms (Jetson)	Yes	Medium
Custom YOLOv8 [44]	NVIDIA A800 (train + inference)	≈ 15.5 ms	No	High
Ours	NVIDIA RTX 2080 Ti (train + inference) Raspberry Pi 5 (inference)	≈ 60 ms (Desktop) ≈ 400 ms (Raspberry)	Yes	Low

[1] Inference time per frame, estimated by 1000 / FPS.
[2] Whether the proposed approach has been implemented in an embedded system.
[3] Monetary cost in case of practical implementation. NVIDIA graphics cards are assumed to be part of a possible server.

the framework is scalable and can be adapted by balancing resolution, coverage, and processing resources to maintain accuracy within this temporal margin.

Overall, the proposed method fills a crucial gap in the state of the art by providing a practical approach for deployment in resource-constrained environments. Unlike previous work that relies on high-power GPUs or fire and smoke analysis based on processed RGB or IR images, this method effectively leverages raw LWIR images, a lightweight detection process, and the feasibility of integrated deployment, making it suitable for early wildfire detection in challenging environments.

Future research should focus on improving accuracy and overcoming some limitations by integrating portable weather stations for real-time recording of environmental factors, such as atmospheric distortions, ambient temperature, relative humidity, wind speed, and various background conditions, to

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2025.3644384

enhance robustness in different early fire detection scenarios. Additionally, optimizing the framework for edge applications and exploring its scalability for large-area monitoring will be crucial to extending its applicability in real-world wildfire detection and prevention systems.

ACKNOWLEDGMENTS

The authors would like to thank Juan Domingo Santana Urbín for his contribution during this research.

REFERENCES

[1] A. J. Sanchez-Fernandez, L. F. Romero, G. Bandera, and S. Tabik, “Vpp: Visibility-based path planning heuristic for monitoring large regions of complex terrain using a uav onboard camera,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 944–955, 2022.

[2] A. Saleh, M. A. Zulkifley, H. H. Harun, F. Gaudreault, I. Davison, and M. Spraggon, “Forest fire surveillance systems: A review of deep learning methods,” *Heliyon*, vol. 10, no. 1, p. e23127, 2024.

[3] M. J. Wooster, G. J. Roberts, L. Giglio, D. P. Roy, P. H. Freeborn, L. Boschetti, C. Justice, C. Ichoku, W. Schroeder, D. Davies, A. M. Smith, A. Setzer, I. Csiszar, T. Strydom, P. Frost, T. Zhang, W. Xu, M. C. de Jong, J. M. Johnston, L. Ellison, K. Vadrevu, A. M. Sparks, H. Nguyen, J. McCarty, V. Tanpipat, C. Schmidt, and J. San-Miguel-Ayanz, “Satellite remote sensing of active fires: History and current status, applications and future requirements,” *Remote Sensing of Environment*, vol. 267, p. 112694, 2021.

[4] Z. Zheng, H. Hu, W. Huang, F. Zhou, Y. Ma, Q. Liu, L. Jiang, and S. Wang, “Wildfire detection based on the spatiotemporal and spectral features of himawari-8 data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.

[5] D. Jurczyński and P. Buchwald, “Mechanisms for data acquisition to train artificial intelligence models for detecting increased susceptibility to fire situations by using internet of things devices and satellite systems,” *SAFETY & FIRE TECHNOLOGY*, vol. 63, pp. 34–43, 07 2024.

[6] M. Sadi, Y. Zhang, W.-F. Xie, and F. M. A. Hossain, “Forest fire detection and localization using thermal and visual cameras,” in *2021 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2021, pp. 744–749.

[7] T. Lewicki and K. Liu, “Multimodal wildfire surveillance with uav,” in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 1–6.

[8] M. A. Radi, P. Li, S. Boumaraf, J. Dias, N. Werghi, H. Karki, and S. Javed, “Ai-enhanced gas flares remote sensing and visual inspection: Trends and challenges,” *IEEE Access*, vol. 12, pp. 56 249–56 274, 2024.

[9] J. A. Sobrino, F. Del Frate, M. Drusch, J. C. Jiménez-Muñoz, P. Manunta, and A. Regan, “Review of thermal infrared applications and requirements for future high-resolution sensors,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2963–2972, 2016.

[10] A. Shamsoshoara, F. Afghah, A. Razi, L. Zheng, P. Z. Fulé, and E. Blasch, “Aerial imagery pile burn detection using deep learning: The flame dataset,” *Computer Networks*, vol. 193, p. 108001, 2021.

[11] X. Chen, B. Hopkins, H. Wang, L. O’Neill, F. Afghah, A. Razi, P. Fulé, J. Coen, E. Rowell, and A. Watts, “Wildland fire detection and monitoring using a drone-collected rgb/ir image dataset,” *IEEE Access*, vol. 10, pp. 121 301–121 317, 2022.

[12] X. Rui, Z. Li, X. Zhang, Z. Li, and W. Song, “A rgb-thermal based adaptive modality learning network for day–night wildfire identification,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 125, p. 103554, 2023.

[13] A. Jong, M. Yu, D. Dhrafani, S. Kailas, B. Moon, K. Sycara, and S. Scherer, “Wit-uas: A wildland-fire infrared thermal dataset to detect crew assets from aerial views,” in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 11 464–11 471.

[14] U. Meleti and A. Razi, “Obscured wildfire flame detection by temporal analysis of smoke patterns captured by unmanned aerial systems,” 2023.

[15] Y. Gao and P. L. Cheng, “Full-scale video-based detection of smoke from forest fires combining vibe and mser algorithms,” *Fire Technology*, vol. 57, pp. 1637 – 1666, 2021.

[16] C. Jin, T. Wang, N. Alhusaini, S. Zhao, H. Liu, K. Xu, and J. Zhang, “Video fire detection methods based on deep learning: Datasets, methods, and future directions,” *Fire*, vol. 6, no. 8, 2023.

[17] M. Fahim-Ul-Islam, N. Tabassum, A. Chakrabarty, S. M. Aziz, M. Shirmohammadi, N. Khonsari, H.-H. Kwon, and M. J. Piran, “Wildfire detection powered by involutional neural network and multitask learning with dark channel prior technique,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 19 095–19 114, 2024.

[18] M. Krišto, M. Ivasic-Kos, and M. Pobar, “Thermal object detection in difficult weather conditions using yolo,” *IEEE Access*, vol. 8, pp. 125 459–125 476, 2020.

[19] S. Jin, T. Wang, H. Huang, X. Zheng, T. Li, and Z. Guo, “A self-adaptive wildfire detection algorithm by fusing physical and deep learning schemes,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 127, p. 103671, 2024.

[20] G. Wang, H. Li, S. Ye, H. Zhao, H. Ding, and S. Xie, “Rfwnet: A multiscale remote sensing forest wildfire detection network with digital twinning, adaptive spatial aggregation, and dynamic sparse features,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–23, 2024.

[21] M. Zhao, W. Li, L. Li, J. Hu, P. Ma, and R. Tao, “Single-frame infrared small-target detection: A survey,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 2, pp. 87–119, 2022.

[22] R. Kou, C. Wang, Q. Fu, Y. Yu, and D. Zhang, “Infrared small target detection based on the improved density peak global search and human visual local contrast mechanism,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 6144–6157, 2022.

[23] H. Liu, H. Shen, W. Ci, Z. Qi, C. Zhou, J. Yao, Q. Shen, and C. Liu, “Effects of the distance and test angle on the precision of infrared temperature measurement,” *IOP Conference Series: Earth and Environmental Science*, vol. 983, no. 1, p. 012025, feb 2022.

[24] T. Wang, Y. Wang, F. Zhao, H. Feng, J. Liu, L. Zhang, N. Zhang, G. Yuan, and D. Wang, “A spatio-temporal temperature-based thresholding algorithm for underground coal fire detection with satellite thermal infrared and radar remote sensing,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 110, p. 102805, 2022.

[25] A. Zhdanova, R. Volkov, A. Sviridenko, G. Kuznetsov, and P. Strizhak, “Influence of compartment fire behavior at ignition and combustion development stages on the operation of fire detectors,” *Fire*, vol. 5, no. 3, 2022.

[26] S. Wang, M. Thomsen, X. Huang, and C. Fernandez-Pello, “Spot ignition of a wildland fire and its transition to propagation,” *International Journal of Wildland Fire*, vol. 33, 06 2024.

[27] M. Storey, M. Almeida, C. Ribeiro, R. Bradstock, and J. Sharples, “Experiments on the influence of spot fire and topography interaction on fire rate of spread,” *PLOS ONE*, vol. 16, p. e0245132, 01 2021.

[28] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics YOLO,” 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[29] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path aggregation network for instance segmentation,” *CoRR*, vol. abs/1803.01534, 2018.

[30] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, “Detrs beat yolos on real-time object detection,” 2024.

[31] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” 2016.

[32] A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, “Searching for mobilenetv3,” *CoRR*, vol. abs/1905.02244, 2019.

[33] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” 2017.

[34] FLIR, “Free teledyne flir thermal dataset for algorithm training,” https://www.flir.eu/oem/adas/adas-dataset-form/, 2023.

[35] J. Liu, X. Fan, Z. Huang, G. Wu, R. Liu, W. Zhong, and Z. Luo, “Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5802–5811.

[36] P. Zippenfenig, “Open-meteo.com weather api,” 2023. [Online]. Available: https://open-meteo.com/

[37] D. Satoh, Y. Takano, R. Sudo, and T. Mochida, “Reduction of communication demand under disaster congestion using control to change human communication behavior without direct restriction,” *Computer Networks*, vol. 134, pp. 105–115, 2018.

[38] J. Sander, A. Cohen, V. R. Dasari, B. Venable, and B. Jalaian, “On accelerating edge ai: Optimizing resource-constrained environments,” 2025.

[39] Wahyono, A. Harjoko, A. Dharmawan, F. D. Adhinata, G. Kosala, and K.-H. Jo, “Real-time forest fire detection framework based on artificial

intelligence using color probability model and motion feature analysis," *Fire*, vol. 5, no. 1, 2022.

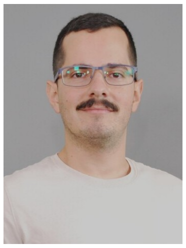
[40] S. Muksimova, S. Umirzakova, S. Mardieva, M. Abdullaev, and Y. I. Cho, "Revolutionizing wildfire detection through uav-driven fire monitoring with a transformer-based approach," *Fire*, vol. 7, no. 12, 2024.

[41] A. Abdusalomov, S. Umirzakova, M. Bakhtiyor Shukhratovich, M. Mukhiddinov, A. Kakhorov, A. Buriboev, and H. S. Jeon, "Drone-based wildfire detection with multi-sensor integration," *Remote Sensing*, vol. 16, no. 24, 2024.

[42] R. Ghali and M. A. Akhloufi, "Deep learning approach for wildland fire recognition using rgb and thermal infrared aerial image," *Fire*, vol. 7, no. 10, 2024.

[43] M. Shahid, S.-F. Chen, Y.-L. Hsu, Y.-Y. Chen, Y.-L. Chen, and K.-L. Hua, "Forest fire segmentation via temporal transformer from aerial images," *Forests*, vol. 14, p. 563, 03 2023.

[44] W. Zhu, S. Niu, J. Yue, and Y. Zhou, "Multiscale wildfire and smoke detection in complex drone forest environments based on yolov8," *Scientific Reports*, vol. 15, 01 2025.



Antonio Galván-Hernández was born in Las Palmas, Spain, in 1998. He received the bachelor's degree in technologies for telecommunications engineering and the M.Sc. in telecommunications engineering from the Universidad de Las Palmas de Gran Canaria (ULPGC), Spain, in 2020 and 2022, respectively. He is currently working on his Ph.D. thesis about algorithms for reducing false alarms in early fire situations. D. Galván-Herández is a member of the Institute for Technological Development and Innovation in Communications (IDeTIC-ULPGC).

Víctor Araña-Pulido (Member, IEEE) was born in Las Palmas, Spain, in 1965. He received the M.Sc. degree from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1990, and the Ph.D. degree from the Universidad de Las Palmas de Gran Canaria (ULPGC), Las Palmas, in 2004. He is currently an Associate Professor and a member of the Institute for Technological Development and Innovation in Communications (IDeTIC-ULPGC). His research interests include the nonlinear design of microwave circuits and common. subsystem units.

Francisco Cabrera-Almeida (Member, IEEE) was born in Las Palmas, Spain, in 1970. He received his M.Sc. and Ph.D. degrees from the Universidad de Las Palmas de Gran Canaria (ULPGC), Las Palmas, in 1997 and 2012, respectively. He is currently an Assistant Professor and a member of the Institute for Technological Development and Innovation in Communications (IDeTIC-ULPGC). His current research interests include the design of communications systems and radio wave propagation.

Pedro Quintana-Morales was born in Santa Cruz de Tenerife, Spain, in 1964. He received the M.Sc. degree from the Universidad Politécnica de Madrid, (UPM) Madrid, Spain, in 1989, and the Ph.D. degree from the Universidad de Las Palmas de Gran Canaria (ULPGC), Las Palmas, Spain, in 2016. He is currently an Assistant Professor with the Signal and Communication Department and a member of the Institute for Technological Development and Innovation in Communications (IDeTIC), ULPGC. His current research interests include signal processing, and data analysis applied to speech, images, biosignals, and sensor networks.