Engineering Applications of Artificial Intelligence 158 (2025) 111324

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Research paper

Addressing false alarms from high-voltage structures in subpixel fire detection

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

IDeTIC, Universidad de Las Palmas de Gran Canaria, 35001, Spain

A R T I C L E I N F O

Dataset link: AFAS-YOLOv8 (Original data)

Keywords:
False alarm
High-voltage structure
Subpixel fire
Thermal image
You only look once

A B S T R A C T

False alarms in subpixel fire detection often arise when high-voltage structures, such as powerlines or towers near thermographic cameras, emit intense infrared radiation that mimics early fire signals at long distances. This paper proposes the study and statistical analysis of You Only Look Once version 8 (YOLOv8) to detect, segment, and isolate these sources of false alarms. YOLOv8 is trained on the Addressing False Alarm Situations (AFAS) dataset, which includes a variety of Long-Wave Infrared (LWIR) and Near-Infrared (NIR) imagery from both aerial and ground-level perspectives. The model achieves a mean Average Precision (mAP) of 0.784 at an Intersection over Union (IoU) threshold of 0.5. The contribution of this work lies in a detailed statistical analysis of YOLOv8 outputs, introducing, among others, the Empirical Cumulative Distribution Function (ECDF) as a metric to assess the relationship between mask overlap and detection confidence. To evaluate the model's robustness under thermal disturbances, synthetic fires are introduced to simulate changes in the scene. The two-sample Kolmogorov–Smirnov (KS) test compares prediction distributions with and without these anomalies, important to ensure that the model performs reliably over a wide range of scenarios so that the presence of these structures can always be determined and isolated. Finally, an energy retention metric is introduced to quantify the probability that the model's predicted masks obscure at least half of an early fire's energy. In critical cases where the fire appears at 2, 3, and 4 pixels from the segmented structures, these probabilities are approximately 7%, 4%, and 3%, respectively.

1. Introduction

Wildfires pose a global threat that demands effective early detection and accurate monitoring to reduce their devastating consequences. The continuous development of wildfire detection systems and the search for new strategies underscore the urgency of this challenge (Barmpoutis et al., 2020). Thermographic technology, widely integrated into fire monitoring systems, provides significant advantages. Thermographic cameras detect infrared thermal radiation, enabling the early identification of heat sources before visible signs, such as smoke, appear. By focusing on infrared energy rather than visible cues, these cameras improve detection accuracy and reduce false alarms (Valero et al., 2021; Carta et al., 2023).

Detecting early fire formation at kilometer distances requires thermographic systems capable of identifying subpixel-scale heat sources, where the burned area is still small and occupies only a few pixels of the camera. For instance, Fig. 1(a) shows a burned area of 3 m^2 with a temperature nearing 864 °C. At a distance of 1600 m, as shown in Fig. 1(b), this burned area appears in the thermal image with a maximum pixel temperature of only 30.01 °C, even though the surrounding terrain averages 39.43 °C. Such challenges frequently arise in rugged, forested terrains with limited observation angles and a kilometer distance to the incipient fire, underscoring the importance of precise subpixel detection to enable rapid emergency response (Zacharakis and Tsihrintzis, 2023).

A critical challenge in detecting subpixel heat sources is the high prevalence of false alarms caused by structures or objects whose infrared emissions appear similar to those of distant fires due to atmospheric attenuation. High-voltage distribution structures have been strongly linked to increased wildfire risks, particularly in steep, forested areas and adverse weather conditions (Bayani et al., 2023; Jahn et al., 2022). These structures frequently exhibit overheating and thermal anomalies driven by weather factors such as wind, humidity, temperature, and by electrical issues like high currents and insulator deterioration resulting from inadequate maintenance (Bigun et al., 2020;

* Corresponding author.
E-mail addresses: antonio.galvan@ulpgc.es (A. Galván-Hernández), victor.arana@ulpgc.es (V. Araña-Pulido), francisco.cabrera@ulpgc.es (F. Cabrera-Almeida), pedro.quintana@ulpgc.es (P. Quintana-Morales).

https://doi.org/10.1016/j.engappai.2025.111324
Received 12 March 2024; Received in revised form 17 April 2025; Accepted 26 May 2025
Available online 13 June 2025

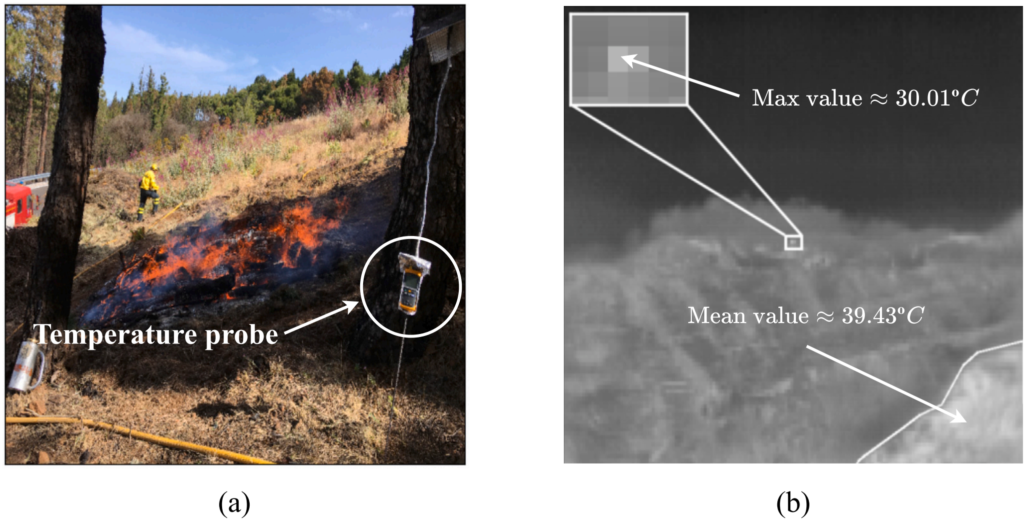


Fig. 1. Near subpixel fire scenario. (a) Prescribed burning carried out by competent authorities. (b) Representation of the burned area in a thermal image of 160 × 120 pixels.

Liu et al., 2020). Moreover, material property changes, such as corrosion or oxide formation on insulators, affect emissivity and infrared radiation emission, misleading thermal readings (Jeong et al., 2023). Consequently, powerlines and towers with altered surface characteristics become significant sources of false alarms, especially during high-temperature, low-wind scenarios (Bigun et al., 2020).

In this context, autonomous mobile detection systems can play a key role by reducing operator deployment time and enhancing efficiency (Perez-Mato et al., 2016). They require simple calibration processes and robust algorithms to swiftly detect potential fires under complex environmental conditions, including extreme heat, adverse weather, and sun reflections. These algorithms must isolate fire signals from non-fire sources to prevent the inclusion of objects whose emitted radiation appears similar to the energy received by the camera from a distant potential fire (Barmpoutis et al., 2020).

This work addresses the challenge of reducing false alarms in subpixel fire detection by focusing on isolating high-voltage structures. The proposed methodology combines computer vision with thermographic cameras to accurately detect and segment specific structures that commonly trigger false alarms, mainly when fires are incipient.

The novelty of the proposed study lies specifically in identifying these structures as significant sources of false alarms in high-depth-of-field scenarios. These scenarios are characterized by multiple infrared radiation sources coexisting at varying distances, with fires potentially igniting kilometers away from the observation point, defining a subpixel scenario. Fig. 2 provides an example of this complex detection environment, where powerlines, towers or other thermal sources may appear prominently within the camera's field of view. The example highlights early fire detection with effective isolation of false alarm sources. A white zoom-in effect marks the detected fire.

To clearly articulate the novelty and contributions of this work, the main points are summarized as follows:

- **Application focus:** This work addresses a problem not mentioned in the state of the art in early fire detection, false alarms caused by the presence of high-voltage structures in thermal images, especially in high depth-of-field scenarios.
- **Dataset contribution:** A new, dedicated dataset is introduced, specifically annotated to include structural elements commonly responsible for false alarms in early fire detection systems based on thermal imagery. The dataset features both aerial and terrestrial views in diverse topographic environments and will be made publicly available.
- **In-depth analysis:** Additional metrics are introduced to evaluate the segmentation model from perspectives not explored in previous work. The analysis focuses on three key aspects:
  - *Confidence-behavior relationship:* Understanding how the model's confidence level correlates with the overlap between predicted and ground truth masks, offering a more intuitive assessment of prediction reliability.
  - *Robustness under subpixel fire scenarios:* Evaluating the model's ability to continue detecting and isolating false-alarm sources during the sudden appearance of thermal disturbances like early fire signals.
  - *Preservation of fire regions:* Measuring how well the model avoids masking true fire areas during incipient fire scenarios, emphasizing the potential risk of obscuring weak thermal signatures.

The remainder of this paper is structured as follows. Section 2 reviews existing false alarm reduction approaches and outlines the research gap addressed in this work. Section 3 introduces the proposed dataset, emphasizing its coverage of high-voltage powerlines and towers. It also details the synthesis of near-subpixel fire anomalies for robustness testing and outlines the workflow adopted in this study. Section 4 presents the experimental outcomes, including both standard metrics and the newly proposed metrics for assessing the model's effectiveness in isolating false-alarm structures while preserving subtle fire signals, as well as a comparison with other approaches to false alarm reduction in the state of the art. Section 5 summarizes the key findings and contributions of this work, particularly the release of a dataset focusing on high-voltage structures, the introduction of new evaluation metrics for subpixel fire scenarios, and a discussion of their implications, along with prospective directions for future research.

2. Related works

In thermography-based fire detection systems, false alarms are often due to solar effects, reflections, hot objects, artificial lighting, and unrelated combustion sources (Barmpoutis et al., 2020). A prominent research direction addresses these issues through signal-based analysis that exploits the unique temporal and spectral patterns of genuine fire events. For instance, Arrue et al. (2000) proposed a real-time infrared–visual system that combines infrared image processing techniques with artificial neural networks and supplementary meteorological data. Their approach, designed to reduce the workload on human operators by filtering out false alarms, leverages information

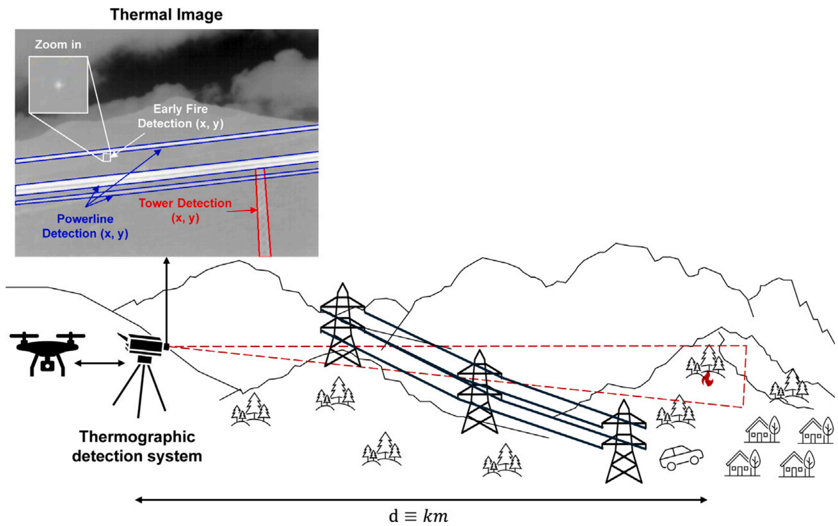


Fig. 2. Representation of a complex subpixel fire scenario where the infrared energy of the powerlines (blue) exceeds the fire energy. The fire has been generated synthetically.

redundancy from dual imaging modalities and integrates a fuzzy expert rule base to support decision-making.

Parallel research has focused on exploiting the spectral characteristics of biomass combustion. Briz et al. (2003) developed algorithms based on the differential behavior of the medium (3–5 μm) and thermal (8–12 μm) infrared spectral regions. Their introduction of the Fire Index (FI) and the complementary Mid-IR Fire Index (MFI) provided a robust framework for discriminating between true fires and false alarms, validated through experimental burns.

Multisensor fusion has emerged as a further enhancement to fire detection systems. Perona et al. (2011) presented an innovative system that fuses radiometric analysis with image, graphical, and motion processing of Long-Wave Infrared (LWIR) and Near-Infrared (NIR) data. Their method, which integrates opto-thermal sensors and adjustable threshold models based on local climatic and orographic factors, substantially improves the reliability of detection by reducing false alarms. Similarly, Georgiades et al. (2019) introduced a multisensor framework that integrates ground-based optical and thermographic cameras with Unmanned Aerial Vehicle (UAV) and environmental sensors. This fusion strategy enables real-time risk assessment and effective early-warning notifications in challenging operational environments. Building on similar fusion concepts, Liu et al. (2023) proposed combining visual and infrared imagery within a single network to reduce both false and missed fire alarms through a custom UAV-captured dataset, their improved model achieved a low false alarm rate and missed alarm rate, highlighting the benefits of multispectral fusion for early forest fire warning.

Addressing deployment challenges further, recent work by Anggreainy et al. (2022) has incorporated fuzzy logic within an Arduino-based data collection framework with different sensors. Their system achieved a high degree of accuracy with a fuzzy output difference, demonstrating promising potential to reduce false alarms in forest fire detection scenarios. In a related vein, Sawant (2024) introduced an integrated detection system that combines smoke sensors, thermal measurements, and image analysis. This multi-model approach reduces false alarms and improves real-time detection capabilities, though concerns regarding cost, maintenance, and data privacy remain.

While the state of the art in false alarm reduction using LWIR and NIR imaging has made significant strides, a considerable research gap persists. Existing approaches have not adequately addressed false alarms caused by high-voltage structures, objects with regular shapes, and high temperatures, or surfaces that reflect sunlight or emit artificial light. The identification and segmentation of these entities using thermographic cameras are areas that have not been fully developed. Moreover, none of these studies have presented scenarios involving multiple infrared radiation sources at varying distances where fires can ignite kilometers away from the observation point. This absence highlights the need for new methodologies capable of handling these complex environments.

3. Material and methods

3.1. Datasets

Two datasets were developed and used in this study. The first, referred to as Addressing False Alarm Situations (AFAS), was created for training and evaluating the detection and segmentation of high-voltage structures such as powerlines and towers. The second is a synthetic extension of AFAS, specifically designed to evaluate the model's robustness in segmenting false alarms and potential fire masking in subpixel fire scenarios.

3.1.1. AFAS dataset

In this work, different datasets for the detection and segmentation of powerlines and towers through the analysis of LWIR and NIR images have been analyzed. Three datasets were selected: the Advanced Driver Assistance Systems (ADAS) dataset by FLIR (2023); the M3DF dataset discussed in the work carried out by Liu et al. (2022); and the Powerline Image Dataset (PID), developed in collaboration with the Turkish Electricity Transmission Company for powerline and tower inspection (Yetgin and Gerek, 2019). A filtering process was employed for the datasets, excluding images with similar scenarios to prevent training bias. Notably, the ADAS and M3DF datasets predominantly showcase urban scenarios, differing from those in this study.

To improve the model's generalization capabilities and address the limitations of existing urban-focused datasets, a targeted measurement campaign was carried out in rural areas. Thermal imagery was collected across diverse landscapes, including mountainous terrain, ravines, and dense vegetation, under varying lighting conditions and weather patterns. These environments closely resemble the deployment conditions of early fire detection systems, where false alarms may originate from high-voltage structural elements. The resulting dataset, named AFAS, also contains thermal images of telephone lines and wooden utility towers. Notably, during data collection, it was observed that the black-coated telephone lines, due to their high emissivity ($\epsilon > 0.95$), exhibit radiometric values significantly above the background under elevated temperatures. A similar behavior was noted in the wooden towers, which also have high emissivity ($\epsilon > 0.8$) according to the study carried out by Pitarma and Crisóstomo (2019).

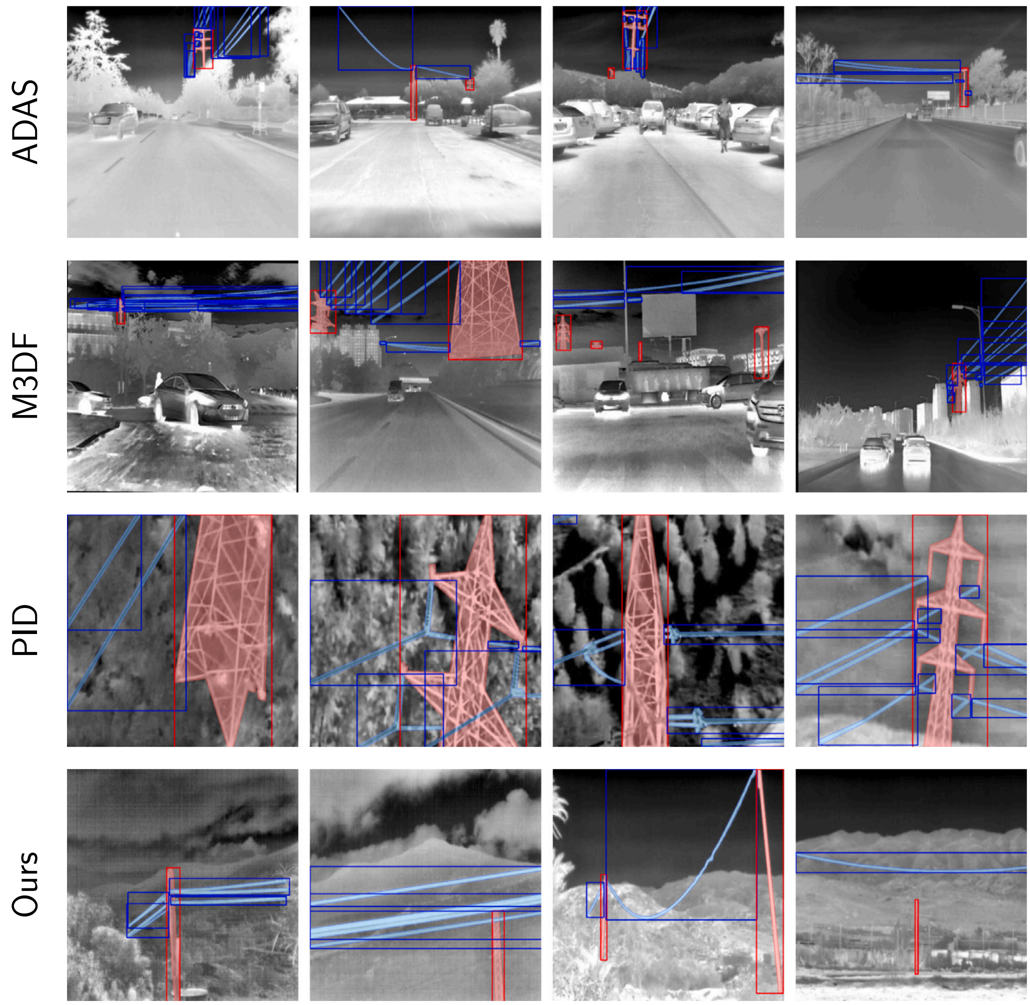


Fig. 3. Different samples of the annotated AFAS dataset.

Table 1
General information about the data used to generate AFAS dataset.

Dataset	Image pairs	Resolution	Camera	Type	Bit depth
ADAS	852	640 × 512	FLIR Tau 2	LWIR	16-bit
M3FD	240	640 × 512	Not specified	LWIR	8-bit
PID	125	576 × 325	Not specified	NIR	8-bit
Ours	**260**	**336 × 256**	**FLIR Tau 2**	**LWIR**	**16-bit**

The details about the thermographic camera used in this work and general information about all the datasets used to create the AFAS dataset are presented in Table 1.

The preprocessing involved normalizing the raw 16-bit LWIR images to an 8-bit format through a linear transformation to maintain a uniform dynamic range across all datasets. Subsequently, Contrast-Limited Adaptive Histogram Equalization (CLAHE) was applied to all images to enhance contrast uniformly before feeding them into the model. Additionally, all images were resized to a standardized resolution of 640 × 640 pixels, ensuring consistency and minimizing the scale variance on detection accuracy.

The AFAS dataset images were manually labeled using the Roboflow labeling tool, producing polygonal annotations specifically for the powerline and tower classes (Dwyer et al., 2024). Examples of the processed dataset and corresponding annotations are illustrated in Fig. 3. For representation purposes, both polygonal contours and bounding boxes were used to show the different objects in the dataset. The powerline class is highlighted in blue, while the tower class is marked in red, providing a visual distinction between the two classes.

3.1.2. Synthetic AFAS dataset

To generate the synthetic dataset, simulated fire outbreaks are introduced into the validation images using a two-dimensional Gaussian distribution, mathematically described as:

$$G(x, y; \mu_x, \mu_y, \sigma_x, \sigma_y) = e^{\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}} \tag{1}$$

where x and y are the coordinates, μ_x and μ_y are the means in the x and y directions, σ_x and σ_x are the standard deviations in the x and y directions.

The new validation image ($I'(x, y)$) is the result of the convolution between the Gaussian effect and the original image ($I(x, y)$), calculated as follows:

$$I'(x, y) = I(x, y) * G(x, y; \mu_x, \mu_y, \sigma_x, \sigma_y) \tag{2}$$

This approach is inspired by observations made during prescribed burn operations, where small, distant fires appeared as smooth, clustered regions of low contrast in thermal imagery as shown in Figs. 1 and 4. The use of this simulation allows for systematic generation of early fire conditions that are representative, reproducible, and scalable for robust model evaluation.

In the image generation process, this effect is applied to each validation image, producing 20 synthetic samples per original image. Two variants of the synthetic dataset are created, each representing different fire intensity levels. The first, **AFAS-SubtleFire**, simulates faint and early-stage fire signatures using Gaussian-based heat sources with lower intensity levels, ranging from 20%–50% of the maximum pixel value in

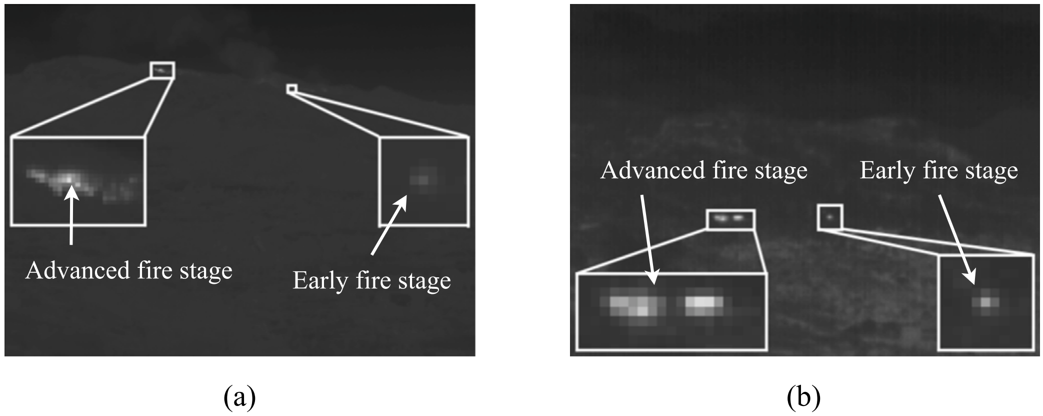


Fig. 4. Different prescribed burnings carried out in collaboration with the competent authorities. (a) Thermal image of 640 × 480 pixels. (b) Thermal image of 200 × 150 pixels.

each image. The second variant, **AFAS-IntenseFire**, introduces more prominent fire-like signals with peaks ranging from 50%–80%. The spatial placement strategy remains consistent in both variants, enabling evaluation of the model's sensitivity and robustness across a range of thermal contrasts. This process has been used to generate the synthetic images shown in Figs. 2 and 5.

3.2. Model overview

You Only Look Once version 8 (YOLOv8) was used to detect and segment out powerlines and towers in this work. It is one of the current state-of-the-art models for computer vision tasks, designed for object detection and tracking, image classification, and instance segmentation (Jocher et al., 2023). Composed of two primary components, the YOLOv8 architecture consists of the backbone and the head. The modified CSPDarknet53 architecture is the backbone, incorporating convolutional layers with cross-stage partial connections to enhance information flow between layers. Meanwhile, the head is responsible for predicting bounding boxes, objectivity scores, and class confidences for detected objects within an image. In essence, the backbone extracts features from the input image, and the head uses these features to make predictions about the objects in the image (Terven et al., 2023).

This model incorporates a variant of the U-Net architecture for segmentation tasks and a variant of the EfficientNet architecture for classification purposes in both detection and segmentation tasks (Terven et al., 2023). The network undergoes initial training to predict object location and class. Subsequently, these predictions are employed to generate a segmentation mask, outlining pixel-level boundaries for each object. This approach enables YOLOv8 to perform instance segmentation.

3.3. Evaluation metrics

The model was evaluated using standard object detection and segmentation metrics to establish a baseline performance. To analyze its behavior under different conditions, additional statistical metrics were applied to the predictions, which provided information on the confidence-behavior relationship, robustness, and spatial accuracy of the model.

3.3.1. Established metrics

The well-established metrics include precision, recall, Average Precision (AP), and mean Average Precision (mAP) when the Intersection over Union (IoU) is set to 0.5. These metrics help to establish a baseline performance reference for the model, ensuring reproducibility and facilitating comparisons with future research.

- **IoU:** For segmentation tasks, the IoU indicates the overlap of the predicted mask with the ground truth mask. This overlap is calculated by measuring the similarity between the two masks as follows:

$$\mathrm{IoU} = \frac{\mathrm{X} \cap \mathrm{Y}}{\mathrm{X} \cup \mathrm{Y}} \tag{3}$$

where $X \cap Y$ is the area where the predicted mask and ground truth mask overlap, while $X \cup Y$ denotes the total area covered by both regions, including both the overlapping and non-overlapping regions.
- **Precision:** The precision is the ratio of correctly predicted positive targets among the total targets predicted as positive. It is calculated as follows:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{4}$$

where TP denotes true positives and FP false positives.
- **Recall:** The recall is the ratio of correctly predicted positive targets among the total actual positive targets. It is calculated as follows:

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{5}$$

where FN stands for false negatives.
- **AP:** The AP serves as a consolidated measure of the precision–recall curve. The higher it is, the better the relationship between precision and recall at different confidence thresholds. It is calculated as follows:

$$AP = \int_{r=0}^{1} P(r)\, dr \tag{6}$$

where $P(r)$ is the precision at the recall level r.
- **mAP:** The mAP is obtained by calculating the average AP values for different classes of objects. When the IoU is set to 0.5, this metric is also known as mAP@50. The mAP metric is calculated as follows:

$$mAP = \frac{1}{n} \sum_{i=1}^{n} AP_i \tag{7}$$

where n is the total number of classes considered.

The precision, recall, AP, and mAP@50 results reported in this work were obtained using the Ultralytics framework, which provides built-in evaluation routines consistent with COCO-style metrics (Jocher et al., 2023).

3.3.2. Additional proposed metrics

In addition to the well-established metrics, the following specialized metrics are introduced:

- **Empirical Cumulative Distribution Function (ECDF):** The ECDF is used to analyze the overlaps between the predicted and ground truth masks and the confidence values of predicted bounding boxes. It is defined as follows:

$$F(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x) \tag{8}$$

where x is the resulting overlap and the detection confidence, $I(x_i \leq x)$ is an indicator function that equals 1 if $x_i \leq x$ and 0 otherwise.
- **Two-sample Kolmogorov–Smirnov (KS) test:** The two-sample KS test evaluates whether two sets of samples arise from the same distribution. In this work the test is used to compare model predictions on AFAS and synthetic AFAS images. This test is non-parametric and does not assume normality or equal variances, making it well-suited for evaluating model predictions that may exhibit non-Gaussian behavior or distributional shifts under perturbations. Its ability to compare the entire shape of two distributions provides a comprehensive view of how model outputs are affected by simulated early fire scenarios (Dodge, 2008). The key ideas behind the two-sample KS test are:
  - **Null hypothesis** (H_0)**:** Both sample sets come from the same distribution.
  - **Test statistic** (D)**:** Measures the maximum vertical distance between the two ECDFs. Larger values indicate greater dissimilarity between the distributions.
  - **p-value:** This parameter represents the probability of observing results at least as extreme as those obtained, assuming that H_0 is true. A small p-value indicates that such results are unlikely to occur by random chance under H_0, providing evidence to reject the null hypothesis. Conversely, a large p-value suggests that the observed differences could reasonably be attributed to chance, supporting H_0. In practice, the null hypothesis is rejected when the p-value falls below a predefined significance threshold, typically $\alpha = 0.05$ (Fisher, 1992; Dodge, 2008).

- **Energy retention:** To assess the extent to which the model masks early fires near high-voltage structures, a distance-based strategy is applied in which synthetic fire effects are introduced at specific distances from known object polygons.
  - Define M as the fraction of synthetic-fire pixels that fall within the predicted mask. It can be calculated as follows:

$$M = \frac{\sum_{i,j} \text{pixels}_{i,j}}{\sum \text{pixels}} \tag{9}$$

  where $\sum_{i,j} \text{pixels}_{i,j}$ represents the sum of the synthetic fire pixel values within the predicted mask, and $\sum$ pixels represents the sum of the total synthetic fire pixel values.
  - The metric of interest is $P(M \geq 0.5)$. It is defined as the probability that at least half of the total fire intensity is captured within the masks predicted by the model. This is calculated as follows:

$$P(M \geq 0.5) = \frac{\text{len}(\{M_i \mid M_i \geq 0.5\})}{N} \tag{10}$$

  where $\text{len}(\{M_i \mid M_i \geq 0.5\})$ is the number of predicted masks for which $M \geq 0.5$, and N is the total number of evaluated images.

Fig. 5 presents an example from the synthetic AFAS dataset to visually demonstrate the evaluated cases in the energy retention analysis.

In Fig. 5(a), the original prediction made by the model is represented. In contrast, Fig. 5(b), (c), and (d) illustrate the different possibilities considered in the evaluation process.

3.4. Experimental procedure

The complete workflow carried out in this work is presented, covering the training process, the inference process and the evaluation process using standard and newly proposed metrics.

3.4.1. Training process

For the training process, all images from the AFAS dataset have been used, providing diverse visual contexts essential for comprehensive model training. The dataset is split into a 70%–30% ratio for training and validation. Furthermore, the data augmentation was exclusively and randomly applied to the training set, incorporating vertical and horizontal flips, static cropping (0% to 30%), Gaussian blur ($\sigma = 0$ to 0.75), random salt and pepper noise injection (1% of pixels replaced), and random brightness changes (−10% to 10%). The final dataset used to train and evaluate the model consists of 3561 images from different scenarios.

The training platform is based on a system running Ubuntu 22.04, equipped with an RTX 2080Ti GPU, an Intel i9-9900X CPU@3.50 GHz, and 32 GB of RAM. The images were resized to 640 × 640 pixels while the annotations were normalized to these dimensions for the training process. As part of this research, the pre-trained M model of YOLOv8 was used as a starting point. The fine-tuning was carried out with the proposed augmented dataset over 300 epochs with default hyperparameter configurations provided by Ultralytics (Jocher et al., 2023).

3.4.2. Inference process

Fig. 6(a) shows a simplified YOLOv8 architecture that makes inference possible, summarizing the key components and data flow. The architecture begins with an input image that is processed by the Backbone (I and II), which features Convolution (Conv) layers, Cross Stage Partial Bottleneck with 2 convolutions (C2f) modules, and Spatial Pyramid Pooling Fusion (SPPF) modules to extract multi-scale features. These modules work together to capture both local details and global context. The extracted features, represented as multi-scale maps (P3, P4, P5), are then routed to the Segmentation head (I and II), which employs a Proto module to generate a set of shared mask prototypes and CV4 modules to produce the corresponding mask coefficients. These components work together to reconstruct precise instance segmentation masks for each detected object. Finally, the network outputs bounding boxes, class indices, class confidences, and segmentation masks, which are used for the evaluation metrics considered (Jocher et al., 2023).

Once trained, the model is used to perform inference on both the original AFAS validation dataset and the synthetic datasets, AFAS-SubtleFire and AFAS-IntenseFire, producing bounding boxes, class labels, confidence scores, and segmentation masks for each. The output predictions from the synthetic datasets are stored separately for comparative analysis.

3.4.3. Evaluation process

Fig. 6(b) illustrates how YOLOv8 predictions are used to compute the previously discussed evaluation metrics, each relying on specific inputs and datasets. Classical segmentation metrics, including IoU, precision, recall, AP, and mAP, are calculated using the AFAS dataset, based on class indices, predicted segmentation masks, and ground truth annotations. The ECDF is also derived from the AFAS dataset, leveraging the relationship between detection confidence scores and mask overlaps, providing insights into the distribution of performance across varying thresholds. The two-sample KS test is applied to prediction outputs from both the AFAS and the synthetic AFAS-SubtleFire and AFAS-IntenseFire datasets, allowing for a statistical comparison of the model's behavior before and after the introduction of simulated fire anomalies. Finally, the energy retention metric is computed using only the AFAS-SubtleFire and AFAS-IntenseFire datasets, measuring the proportion of fire-related energy captured within the predicted masks to assess how early fire signals might be obscured or detected.

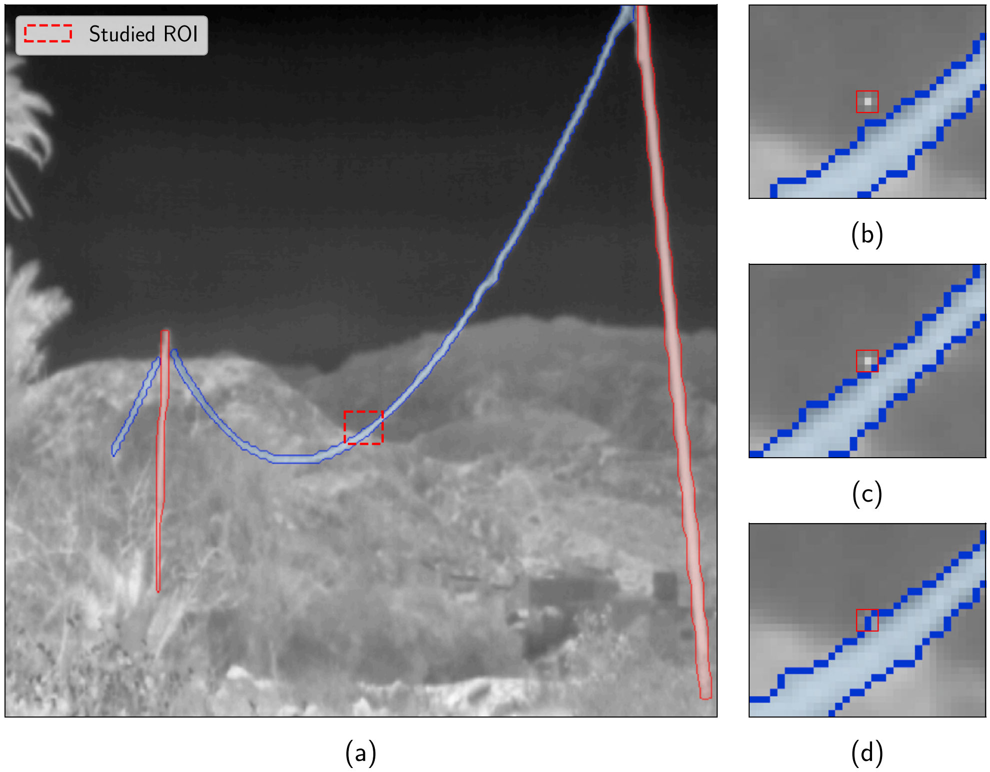


Fig. 5. Samples from synthetic AFAS dataset. (a) Original prediction. (b) Case where the fire is not masked. (c) Case where the fire is partially masked. (d) Case where the fire is almost completely masked.

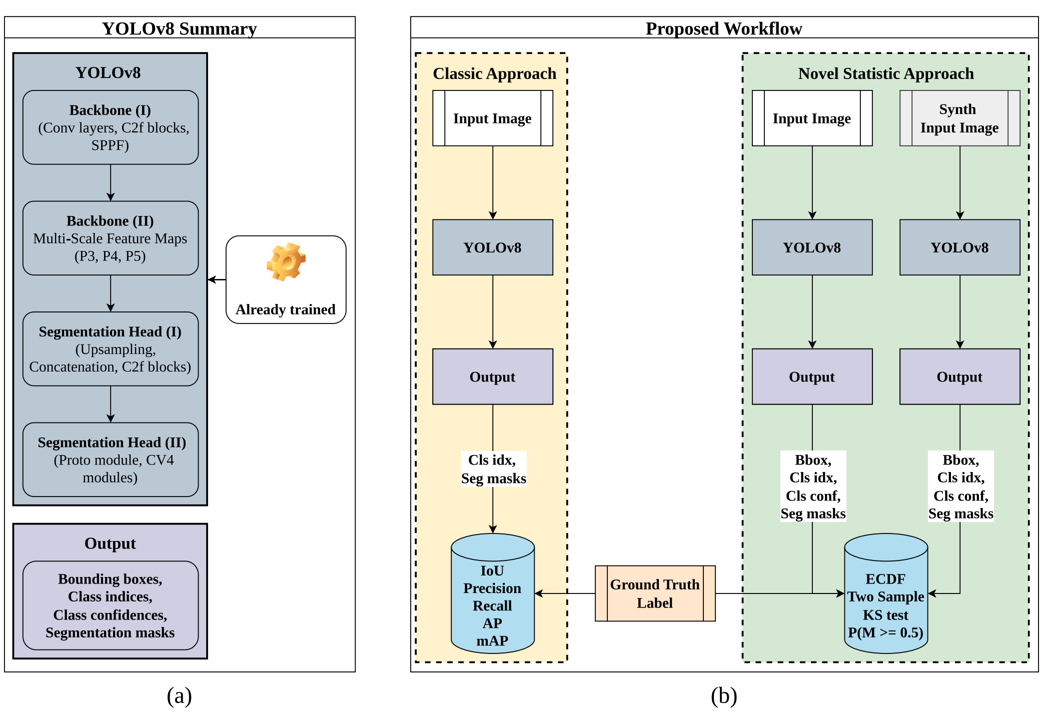


Fig. 6. Methodology block diagram. (a) YOLOv8 simplified architecture and output predictions. (b) Proposed workflow followed in this work.

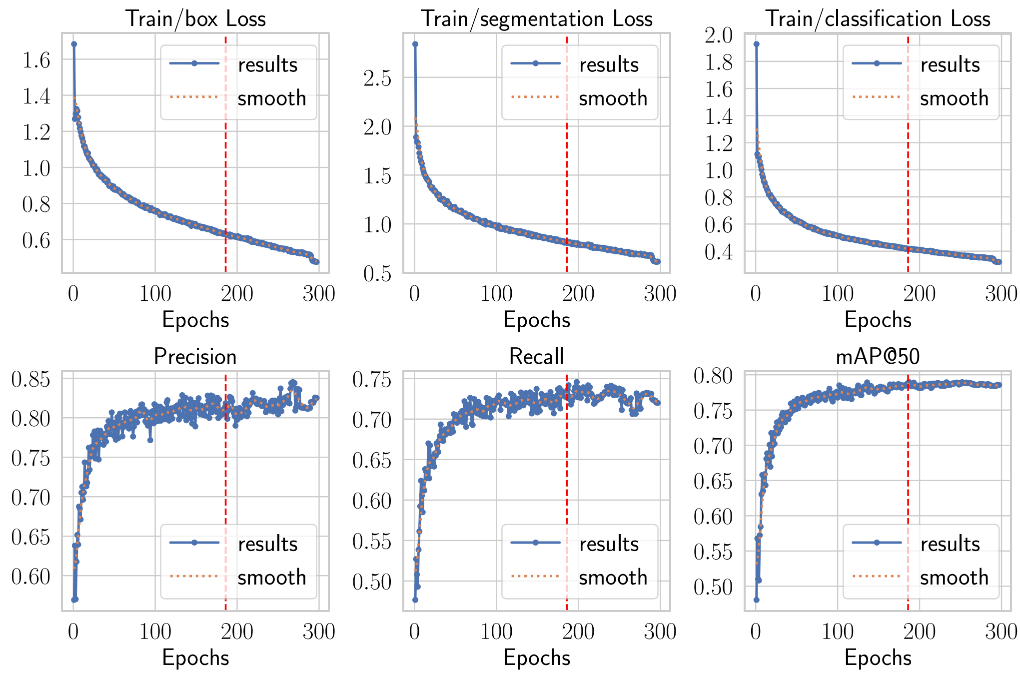


Fig. 7. Training process results over 300 epochs.

Table 2
Evaluation metrics results.

Class	Precision	Recall	AP	mAP@50	Inference time
Powerline	0.808	0.71	0.762	0.784	10.2 ms
Tower	0.83	0.743	0.807		

4. Results and discussion

4.1. Training results

The curves derived from the training process are shown in Fig. 7. The graph illustrates the box loss, segmentation loss, and class loss for the augmented training set, along with the same representation for recall, precision, and mAP@50 for segmentation tasks. The final model used in this study reaches its maximum value of mAP@50 at epoch 186.

4.2. Comparative analysis of detection and segmentation tasks

The evaluation process consisted of applying the established metrics to the predictions made by YOLOv8 on the AFAS validation dataset. The results of this evaluation process are shown in Table 2, where inference time is also displayed in milliseconds (ms).

Several existing works, such as those carried out by Abdelfattah et al. (2023) and by Yang et al. (2022), have explored powerline and tower segmentation using different methodologies and datasets. The study conducted by Abdelfattah et al. (2023) compares different deep learning methods for segmenting powerlines and towers in the visible range. Their proposed method relies on Generative Adversarial Networks (GANs) for segmentation and achieves a precision value of **0.863**. However, compared with an architecture based on UNET++, the latter achieves a maximum recall of **0.591**. The study presented by Yang et al. (2022) evaluates an attention-based segmentation method within an encoder–decoder framework and uses images from the PID dataset, which is also used in the present work. The study achieves a precision value of **0.852** for visible images and **0.856** for infrared images. Additionally, the AP values are reported as **0.910** for visible images and **0.929** for infrared images.

Although this paper refers to those studies and results, it is important to highlight the differences in the experimental setups. For instance, Abdelfattah et al. (2023) used a dataset of over 1100 UAV-captured images in the visible range, with a specific focus on close-range scenarios. On the other hand, Yang et al. (2022) employed the PID dataset for their evaluation, featuring 200 images with particular features extracted from the near-infrared spectrum. To the best of the author's knowledge, no open dataset in the thermal range has been proposed specifically for the study of high-voltage powerlines and towers. The presented dataset, comprising 1477 images representing different scenarios, including both aerial and terrestrial views, aims to fill this gap and will be fully available for future research.

Rather than positioning Table 2 results in direct competition with those of prior studies, they are presented to provide contextual background and to underscore the diversity of existing approaches for high-voltage structures segmentation. Moreover, these metrics help establish a baseline performance reference for the model, promoting reproducibility and facilitating meaningful comparisons with subsequent research in similar domains.

The main contribution of this work lies in the introduction of several statistical evaluation metrics applied to the model predictions. While these metrics are demonstrated using the widely adopted YOLOv8 architecture, they are inherently model-independent and intended to enable consistent, interpretable evaluation across future segmentation models and datasets.

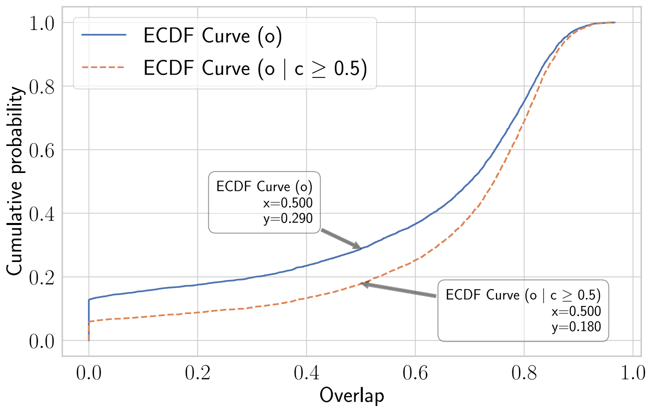
4.3. In-depth analysis of model predictions

The results discussed in this section were obtained for a minimum confidence threshold of 0.1 and an IoU value of 0.5.

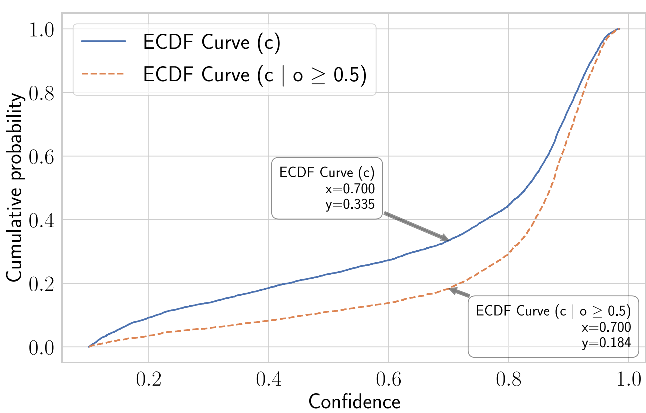
4.3.1. Confidence-behavior relationship

Building on the ECDF-based evaluation, this subsection explores how the model's spatial localization performance correlates with its confidence scores. The analysis focuses on understanding the distribution of overlaps and detection confidence, offering insight into when and how the model's predictions can be considered reliable.

The resulting ECDF curves of this work are shown in Fig. 8.



(a)



(b)

Fig. 8. ECDF analysis. (a) Overlap curves: one for all detections and one for detections with confidence greater than 0.5 (denoted as c | o ≥ 0.5), showing the IoU distribution between predicted and ground truth masks. (b) Confidence curves: one for all detections and one for detections with overlap greater than 0.5 (denoted as c | o ≥ 0.5), illustrating the distribution of bounding box confidence scores.

Fig. 8(a) illustrates the model's localization capability using cumulative probability. The blue line represents the frequency with which the model achieves specific overlap levels with the ground truth mask. For instance, at an X-axis value of 0.5, it is revealed that 29% of the samples have an overlap below 0.5, serving as an implicit threshold for identifying false positives. Simultaneously, the orange dashed line focuses on the cumulative probability of overlap for confidence scores exceeding 0.5. A positive correlation emerges, indicating improved localization capability as the confidence level of the bounding box increases. In Fig. 8(b), the data distribution reflects the confidence levels provided by the model. Notably, when focusing on instances where the overlap exceeds 0.5, indicating true positives, it is observed that about 81% of these detections possess a confidence level surpassing 0.7. This implies that the model reliably identifies true positives with a high degree of confidence.

From a practical standpoint, these findings inform how an operator might set the model's confidence threshold depending on the cost of false positives versus missed detections. By leveraging the resulting ECDF plots, decision makers can visually assess the trade-offs between confidence and detection accuracy, allowing them to set an optimal threshold tailored to the specific needs of the application. This approach simplifies the threshold selection process by providing a clear, data-driven tool that could align model behavior with real-world operational constraints.

4.3.2. *Robustness under subpixel fire scenarios*

To evaluate the model's ability to continue isolating potential false-alarm sources during early-stage fire signals, a two-sample KS test was conducted between the original and synthetic AFAS datasets. The goal of this analysis is to understand whether the model's outputs remain statistically consistent when subtle, near-subpixel thermal anomalies are introduced. The test was applied to model predictions across a range of confidence thresholds (c_0), from 0.10 to 0.99 in incremental steps. At each confidence threshold, the distributions of overlaps between predicted and ground truth masks were compared between the original dataset and its fire-augmented variants (AFAS-SubtleFire and AFAS-IntenseFire).

This analysis provides a series of KS statistics (D values) and corresponding p-values, offering insight into how the model's detection behavior shifts under increasingly confident predictions. The results are plotted as curves in Fig. 9, which enhances interpretability by intuitively highlighting confidence ranges in which the model remains robust and those where it may become more sensitive to synthetic perturbations. This visualization provides a better understanding of whether the model can reliably maintain segmentation performance in the presence of thermal changes caused by very early fire signatures, avoiding their obscuration.

In Fig. 9(a) and (b), the test outcomes are shown when applied over different bounding box confidence values obtained from the model. Fig. 9(c) and (d) showcase the corresponding results for cases involving the overlap between predicted and ground truth masks. The X-axis represents the confidence intervals at which the model has been evaluated, and the Y-axis displays the D and p-value results for each case. The blue lines represent the outcomes of the two-sample KS test comparing the original predictions with those of the AFAS-SubtleFire dataset variant. Similarly, the orange dashed lines illustrate the results of the same test for the AFAS-IntenseFire dataset variant.

In cases where the confidence level $c_0 < 0.9$, the KS statistic D remains relatively low across both synthetic variants, indicating minimal divergence in the ECDFs of the prediction scores. This suggests that, within this range, the model's outputs are largely consistent before and after the synthetic perturbation. For confidence levels $c_0 > 0.9$, an increase in the D value is observed. This increase is likely due to the smaller sample sizes at these high confidence levels, which may amplify statistical variation and reduce test stability. These regions are visually indicated by a red patch in Fig. 9 to emphasize the associated uncertainty.

As shown in Fig. 9(b), the p-values associated with the AFAS-IntenseFire dataset variant are particularly low between confidence intervals of 0.10 and 0.16, suggesting a statistically significant difference between the original and perturbed data distributions in this narrow range. This region is marked with an orange patch in the figure. Beyond a confidence value of 0.16, the p-values increase and frequently exceed the common significance threshold ($\alpha = 0.05$), providing less evidence to reject the null hypothesis.

These results reveal that the model exhibits localized sensitivity to thermal changes in low-confidence regions but maintains overall stability at moderate to high confidence levels. The KS test provides preliminary insight into the model's ability to isolate potential false-alarm sources during the sudden appearance of thermal disturbances, such as early fire signals, ensuring that these do not interfere with or obscure the actual fire detection.

Further statistical analyses, including other non-parametric tests, assessments of variance, confidence interval stability, and robustness under diverse environmental conditions, are essential to validate these trends. Real-world testing remains a crucial next step to confirm whether the observed stability generalizes to actual early fire situations.

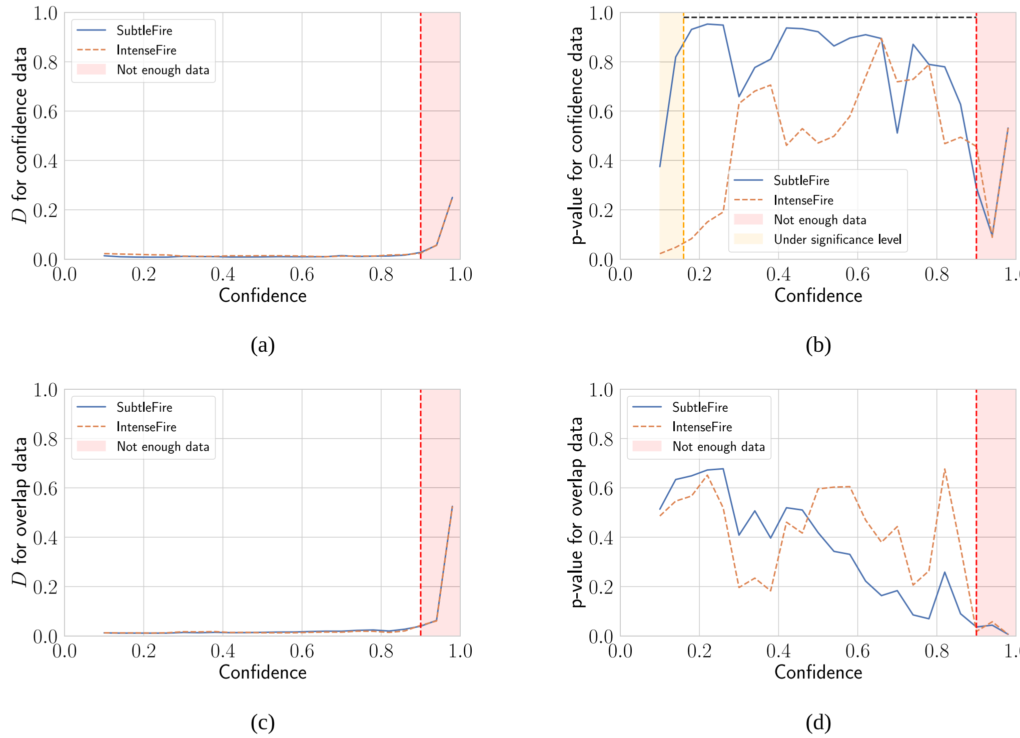


Fig. 9. Two-sample KS test analysis. (a) KS statistic (D value) for confidence samples. (b) Corresponding p-values for confidence samples. (c) KS statistic (D value) for overlap samples. (d) Corresponding p-values for overlap samples.

4.3.3. Preservation of fire regions

Based on the energy retention metric $P(M \geq 0.5)$, this section shows the extent to which the model does not mask the initial early fires by quantifying the fraction of synthetic thermal energy possibly retained within the predicted masks. The proposed metric allows for a probabilistic understanding of how segmentation performance may affect the detection of subtle fire cues in safety-critical contexts.

The evaluation curves for both synthetic dataset variants are shown in Fig. 10. In this analysis, synthetic fires were positioned at varying distances from the nearest segmented object, ranging from 2 to 8 pixels. The results are color-coded according to these distances to illustrate how proximity influences the likelihood of fire energy being masked. Specifically, Fig. 10(a) shows the results for the AFAS-SubtleFire dataset, while Fig. 10(b) corresponds to the AFAS-IntenseFire dataset.

Across both synthetic datasets, the maximum probability that half of the fire energy is masked occurs at the shortest distances, with values approximately 7%, 4%, 3%, and 2% for distances of 2, 3, 4, and 5 pixels, respectively. This trend indicates that as the synthetic fire is placed farther from the segmented object, the likelihood of it being obscured decreases. Additionally, a clear inverse relationship is observed between the model's prediction confidence and the probability of masking, suggesting that high-confidence detections are generally more reliable in not obscuring fire-relevant regions.

The comparison between AFAS-SubtleFire and AFAS-IntenseFire reveals similar probabilities for both variants, indicating that the model's ability to preserve synthetic fire regions remains largely consistent whether the perturbations are subtle or intense. This outcome suggests that, for the scenarios tested, the model does not exhibit a significant bias towards either weaker or stronger thermal anomalies, thereby maintaining comparable performance across both levels of fire intensity.

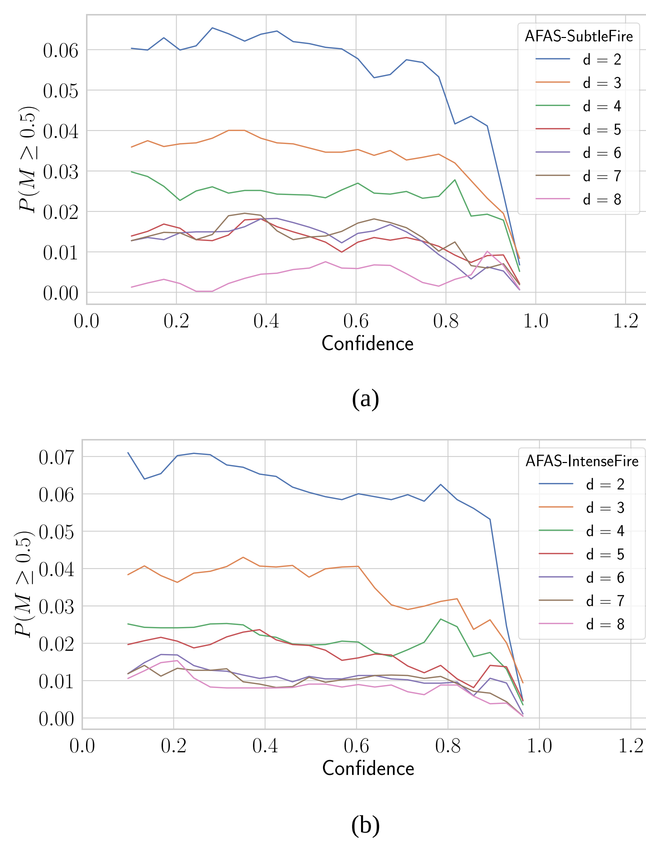


Fig. 10. Mask quality evaluation results. (a) $P(M \geq 0.5)$ results for AFAS-SubtleFire dataset. (b) $P(M \geq 0.5)$ results for AFAS-IntenseFire dataset.

4.4. Qualitative comparison with other false alarm reduction approaches

A variety of strategies for reducing false alarms in wildfire and early fire detection systems have been explored in the literature, each

Table 3
Overview of false alarm reduction approaches studied in the literature (I).

Study	Sensors used	False alarm sources
Arrue et al. (2000)	IR, Visual, Meteorological, Geographic	Sunlight effects, reflections, hot objects, artificial lights, non-wildfire combustion, fog, clouds, vehicles
Briz et al. (2003)	MIR, LWIR	Sunglints, industrial hotspots, sun-heated ground
Perona et al. (2011)	NIR, LWIR, Visual	Hotspots, bright objects, airborne objects mistaken for smoke
Georgiades et al. (2019)	LWIR, Visual	Various potential false positives, not named specifically
Anggreainy et al. (2022)	Smoke, Fire, Temperature	False alarms in general; not specified by type
Liu et al. (2023)	LWIR, Visual	Non-flame heat sources in IR, bright visible objects at night
Sawant (2024)	LWIR, Visual, Smoke	Traditional system errors (smoke or heat), imprecise fire boundary detection
Ours	**LWIR**	**High-voltage structures; powerlines and towers**

Table 4
Overview of false alarm reduction approaches studied in the literature (II).

Study	Approach	Contribution
Arrue et al. (2000)	Thresholding, oscillation, analysis, fuzzy rules	Filters light, heat, and reflection artifacts, via multi-sensor fusion and rule-based logic
Briz et al. (2003)	Spectral analysis (FI, MFI)	Discriminates fire from glints, hotspots, and heated terrain using infrared features
Perona et al. (2011)	Motion and smoke detection	Suppresses false positives from bright and moving objects
Georgiades et al. (2019)	Fusion, risk modeling, UAV-based verification	Minimizes false alerts via UAV confirmation and terrain risk mapping
Anggreainy et al. (2022)	Fuzzy logic with sensors	Accurate detection using fuzzy rules to reject benign heat sources
Liu et al. (2023)	Infrared & Visual fusion with attention mechanism	Reduces false and missed alarms using multispectral enhancement
Sawant (2024)	Detection and segmentation Spatio-temporal analysis	Improves accuracy via visual confirmation and dynamic cues
Ours	**Detection and segmentation In-depth model output analysis**	**Addresses uncertainty, robustness, and fire occlusion effects**

tackling different sources of spurious alerts and employing different sensor modalities. Tables 3 and 4 illustrate the broad spectrum of methods and highlight the diversity of both the sensors used and the sources of false alarms they target.

Despite this variety, Table 3 reveals that few studies directly address high-voltage structures as a significant cause of false alarms in the thermal domain. In many practical settings, such structures appear with high thermal contrast or reflective properties, frequently triggering spurious alerts if not accurately identified and separated from genuine fire signals. This work focuses specifically on this overlooked but present issue.

Compared to traditional threshold-based, fusion-based, or rule-based techniques, the proposed approach emphasizes detection and segmentation coupled with in-depth statistical analyses of the resulting model outputs. Rather than manually tuning thresholds or heuristics, the model's predictions are evaluated using metrics such as the ECDF, two-sample KS tests, and the proposed energy retention criterion. These analyses provide a quantitative, model-independent way of assessing when the segmentation model may obscure early fire signals or become sensitive to synthetic perturbations in the scene. By quantifying both the confidence scores and spatial overlap of predictions, this methodology offers a rigorous view of how high-voltage structures can be isolated without compromising the detection of subtle thermal anomalies.

This method complements existing false alarm reduction techniques by presenting a specialized solution for thermal imagery with high-voltage structures, accompanied by robust evaluation metrics that can be applied to any segmentation architecture. This targeted yet broadly applicable approach aims to improve reliability in safety-critical fire monitoring scenarios, ensuring that early-stage fire cues remain visible and distinguishable despite the presence of common, non-fire thermal objects.

5. Conclusions

This work presents the use of YOLOv8 for detecting and segmenting powerlines and towers, which emit significant infrared radiation under high temperatures or abnormal conditions. These emissions can interfere with thermal imagery, masking early-stage or distant wildfires and increasing the risk of missed detections. The goal is to enhance wildfire detection systems by reducing false alarms caused by such structures, especially when fires can occur at the subpixel level.

A key contribution of this work is the introduction of the AFAS dataset, specifically designed to represent different conditions relevant to fire detection systems. Unlike previous datasets, which often focused on visible or NIR spectrum in close-range scenarios, this dataset includes 1477 LWIR and NIR images capturing a wide range of situations across orographically diverse environments. The trained model achieves a mAP of 0.784 at an IoU threshold of 0.5, establishing a preliminary performance baseline for future improvements and comparative studies.

This work introduces three complementary evaluation methods: the ECDF, the two-sample KS test, and an energy retention-based metric. Together, these approaches provide a deeper understanding of model behavior beyond what standard performance metrics can provide.

The ECDF intuitively visualizes the relationship between mask overlap and bounding box detection confidence, enhancing interpretability and supporting a more nuanced evaluation of segmentation quality.

The two-sample KS test is employed to compare model outputs on original versus synthetically generated datasets. The results indicate that the model exhibits statistically similar behavior across most confidence levels, particularly beyond a threshold of 0.16. This suggests a degree of robustness under simulated early fire scenarios. However, as these findings are based on synthetic data, further validation on real-world events is essential to assess their practical relevance. In addition, complementary statistical analyses, such as other non-parametric methods, should be considered in future work to strengthen and extend the robustness assessment.

The energy retention metric measures the probability that at least half of the energy associated with an early fire lies within the predicted mask. The analysis reveals that the worst-case scenarios occur at detected object distances of 2, 3, and 4 pixels, with corresponding occurrence rates of 7%, 4%, and 3%, respectively. Furthermore, for the scenarios tested, the model shows no significant bias towards weaker or stronger thermal anomalies, thus maintaining comparable performance in both fire intensity ranges.

Future directions for this research include expanding the dataset to encompass a wider range of potential sources of false alarms, such as animals, human activity, reflective surfaces, and vehicle machinery. This expansion will also focus on incorporating more diverse environments, including urban and densely forested environments, to improve the generalizability of the model. Additional efforts will involve testing the model on real-world early fire datasets, if available, and refining preprocessing steps to enhance robustness across imaging sources. Moreover, there is strong potential for exploring real-time implementation and integrating the proposed methodology into operational wildfire detection systems.

CRediT authorship contribution statement

Antonio Galván-Hernández: Writing – review & editing, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Víctor Araña-Pulido:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Francisco Cabrera-Almeida:** Writing – review & editing, Visualization, Supervision, Resources. **Pedro Quintana-Morales:** Writing – review & editing, Software, Formal analysis.

Funding

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, Spain, 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.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Antonio David Galvan Hernandez reports financial support was provided by University of Las Palmas de Gran Canaria. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

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

Data availability

The data is available on the Mendeley data website and on the GitHub of the corresponding author.

AFAS-YOLOv8 (Original data) (Mendeley Data)

References

Abdelfattah, R., Wang, X., Wang, S., 2023. PLGAN: Generative adversarial networks for power-line segmentation in aerial images. IEEE Trans. Image Process. 32, 6248–6259. http://dx.doi.org/10.1109/TIP.2023.3321465.

Anggreainy, M.S., Kurniawan, B., Kurniadi, F.I., 2022. Reduced false alarm for forest fires detection and monitoring using fuzzy logic algorithm. Int. J. Adv. Comput. Sci. Appl. 13 (7), http://dx.doi.org/10.14569/IJACSA.2022.0130764.

Arrue, B., Ollero, A., Matinez de Dios, J., 2000. An intelligent system for false alarm reduction in infrared forest-fire detection. IEEE Intell. Syst. their Appl. 15 (3), 64–73. http://dx.doi.org/10.1109/5254.846287.

Barmpoutis, P., Papaioannou, P., Dimitropoulos, K., Grammalidis, N., 2020. A review on early forest fire detection systems using optical remote sensing. Sensors 20 (22), http://dx.doi.org/10.3390/s20226442.

Bayani, R., Waseem, M., Manshadi, S.D., Davani, H., 2023. Quantifying the risk of wildfire ignition by power lines under extreme weather conditions. IEEE Syst. J. 17 (1), 1024–1034. http://dx.doi.org/10.1109/JSYST.2022.3188300.

Bigun, A.A., Girshin, S.S., Goryunov, V.N., Shepelev, A.O., Pruss, S.Y., Tkachenko, V.A., 2020. Assessment of climatic factors influence on the time to reach maximum wire temperature of overhead power lines. Prz. Elektrotech. R. 96, nr 8, 39–42. http://dx.doi.org/10.15199/48.2020.08.08.

Briz, S., de Castro, A., Aranda, J., Meléndez, J., López, F., 2003. Reduction of false alarm rate in automatic forest fire infrared surveillance systems. Remote Sens. Environ. 86 (1), 19–29. http://dx.doi.org/10.1016/S0034-4257(03)00064-6.

Carta, F., Zidda, C., Putzu, M., Loru, D., Anedda, M., Giusto, D., 2023. Advancements in forest fire prevention: A comprehensive survey. Sensors 23 (14), http://dx.doi.org/10.3390/s23146635.

Dodge, Y., 2008. Kolmogorov–Smirnov test. In: The Concise Encyclopedia of Statistics. Springer New York, New York, NY, pp. 283–287. http://dx.doi.org/10.1007/978-0-387-32833-1_214.

Dwyer, B., Nelson, J., Hansen, T., et al., 2024. Roboflow (version 1.0) [software]. https://roboflow.com. Computer vision.

Fisher, R.A., 1992. Statistical methods for research workers. In: Breakthroughs in Statistics: Methodology and Distribution. Springer New York, New York, NY, pp. 66–70. http://dx.doi.org/10.1007/978-1-4612-4380-9_6.

FLIR, 2023. FREE teledyne FLIR thermal dataset for algorithm training. https://www.flir.eu/oem/adas/adas-dataset-form/.

Georgiades, G., Papageorgiou, X.S., Loizou, S.G., 2019. Integrated forest monitoring system for early fire detection and assessment. In: 2019 6th International Conference on Control, Decision and Information Technologies. CoDIT, pp. 1817–1822. http://dx.doi.org/10.1109/CoDIT.2019.8820548.

Jahn, W., Urban, J.L., Rein, G., 2022. Powerlines and wildfires: Overview, perspectives, and climate change: Could there be more electricity blackouts in the future? IEEE Power Energy Mag. 20 (1), 16–27. http://dx.doi.org/10.1109/MPE.2021.3122755.

Jeong, S., Kim, M.-G., Kim, J.-H., Oh, K.-Y., 2023. Thermal monitoring of live-line power transmission lines with an infrared camera mounted on an unmanned aerial vehicle. Struct. Heal. Monit. 22 (6), 3707–3722. http://dx.doi.org/10.1177/14759217231156359.

Jocher, G., Chaurasia, A., Qiu, J., 2023. Ultralytics YOLOv8. URL https://github.com/ultralytics/ultralytics.

Liu, J., Fan, X., Huang, Z., Wu, G., Liu, R., Zhong, W., Luo, Z., 2022. 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. pp. 5802–5811. http://dx.doi.org/10.1109/CVPR52688.2022.00571.

Liu, X., Miao, X., Jiang, H., Chen, J., 2020. Data analysis in visual power line inspection: An in-depth review of deep learning for component detection and fault diagnosis. Annu. Rev. Control. 50, 253–277. http://dx.doi.org/10.1016/j.arcontrol.2020.09.002.

Liu, Y., Zheng, C., Liu, X., Tian, Y., Zhang, J., Cui, W., 2023. Forest fire monitoring method based on UAV visual and infrared image fusion. Remote. Sens. 15 (12), http://dx.doi.org/10.3390/rs15123173.

Perez-Mato, J., Arana, V., Cabrera-Almeida, F., 2016. Real-time autonomous wildfire monitoring and georeferencing using rapidly deployable mobile units. IEEE Aerosp. Electron. Syst. Mag. 31 (2), 6–15. http://dx.doi.org/10.1109/MAES.2016.150057.

Perona, G., losso, a., Corgnati, L., 2011. False alarm reduction in forest fires detection with low-cost commercial sensors. Eur. J. Remote. Sens. http://dx.doi.org/10.5721/ItJRS20114316.

Pitarma, R., Crisóstomo, J., 2019. An approach method to evaluate wood emissivity. J. Eng. 2019, 1–9. http://dx.doi.org/10.1155/2019/4925056.

Sawant, S., 2024. Integrated fire detection system using ML and IOT. Int. J. Res. Appl. Sci. Eng. Technol. 12, 2091–2100. http://dx.doi.org/10.22214/ijraset.2024.61810.

Terven, J., Córdova-Esparza, D.-M., Romero-González, J.-A., 2023. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Mach. Learn. Knowl. Extr. 5 (4), 1680–1716. http://dx.doi.org/10.3390/make5040083.

Valero, M.M., Verstockt, S., Butler, B., Jimenez, D., Rios, O., Mata, C., Queen, L., Pastor, E., Planas, E., 2021. Thermal infrared video stabilization for aerial monitoring of active wildfires. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 14, 2817–2832. http://dx.doi.org/10.1109/JSTARS.2021.3059054.

Yang, L., Fan, J., Huo, B., Li, E., Liu, Y., 2022. PLE-net: Automatic power line extraction method using deep learning from aerial images. Expert Syst. Appl. 198, 116771. http://dx.doi.org/10.1016/j.eswa.2022.116771.

Yetgin, Ö.E., Gerek, Ö.N., 2019. Powerline image dataset (infrared-IR and visible light-VL). http://dx.doi.org/10.17632/n6wrv4ry6v.8.

Zacharakis, I., Tsihrintzis, V.A., 2023. Integrated wildfire danger models and factors: A review. Sci. Total Environ. 899, 165704. http://dx.doi.org/10.1016/j.scitotenv.2023.165704.