

Article

Not peer-reviewed version

Early Wildfire Smoke Detection with a Multi-Resolution Framework and Two-Stage Classification Pipeline

[Gihwan Jung](#), [Tae-Hyuk Ahn](#), [Byungseok Min](#) *

Posted Date: 13 January 2026

doi: [10.20944/preprints202601.0913.v1](https://doi.org/10.20944/preprints202601.0913.v1)

Keywords: early smoke detection; wildfire monitoring; small object detection; multiresolution



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Early Wildfire Smoke Detection with a Multi-Resolution Framework and Two-Stage Classification Pipeline

Gihwan Jung ¹, Tae-Hyuk Ahn ¹ and Byungseok Min ^{2,*}

¹ Department of Computer Science, Saint Louis University, Saint Louis, MO 63103, USA
² Department of Artificial Intelligence Data Science, Sejong University, Seoul 05006, Republic of Korea
* Correspondence: bmin@sejong.ac.kr

Abstract

Early wildfire smoke detection is critical for preventing small ignitions from escalating into large-scale fires, yet early-stage smoke plumes are often faint, low-contrast, and spatially small. When full-resolution frames are resized to satisfy fixed-input detector architectures and enable efficient batched GPU inference, these subtle cues are further diminished, leading to missed detections and unreliable scores near deployment thresholds. Existing remedies such as multi-scale inference, slicing/tiling, or super-resolution could improve sensitivity, but typically incur substantial overhead from multiple forward passes or added network components, limiting real-time use on resource-constrained platforms. To mitigate these challenges, we propose a composite multi-resolution detection framework that improves sensitivity to small smoke regions while maintaining single-pass inference. Motivated by the fact that most operational wildfire monitoring systems rely on UAV platforms and mountain-top CCTV surveillance, their wide-field imagery typically contains a large sky region above the horizon where early smoke is most likely to first become visible. Accordingly, crop placement is guided by a skyline prior that prioritizes this high-probability sky band while retaining the remaining scene for global context. A dynamic compositing stage stacks a global view with a high-resolution, sky-aligned band into a standard square detector input, preserving context with minimal added cost. Detections from the two views are reconciled via coordinate restoration and non-maximum suppression. For deployment, a lightweight second-stage classifier selectively re-evaluates low-confidence detections to stabilize decisions near a fixed operating threshold without retraining the detector. Across Early Smoke and Pyro-SDIS, the method improves AP@0.5 by +0.060/+0.036 and AP_{small} by +0.151/+0.025, respectively, with accuracy-efficiency trade-offs supported by ablation studies and deployment-time latency analysis.

Keywords: early smoke detection; wildfire monitoring; small object detection; multiresolution

1. Introduction

Early fire detection plays a crucial role in preventing small ignition events from escalating into large-scale wildfires that cause extensive ecological damage, economic loss, and threats to human life and infrastructure [1]. Once a fire grows beyond its incipient stage, suppression becomes significantly more difficult and costly, particularly in remote or mountainous regions where access is limited and response times are long [2]. As climate change and prolonged drought conditions increase both the frequency and intensity of wildfires, the ability to identify fire-related signals at the earliest possible moment has become a central requirement for modern wildfire management systems [3]. In operational settings, many detection systems rely on vision-based monitoring using UAVs equipped with high-resolution RGB cameras or fixed mountain-top surveillance CCTVs [4]. These platforms enable wide-area coverage and continuous observation without the need for dense ground sensor deployment, making single-image smoke detection a common and scalable paradigm [5]. However,

early-stage smoke presents a uniquely challenging visual target: it often occupies only a very small portion of the image—a small fraction of the total area—while exhibiting low contrast, diffuse boundaries, and strong visual similarity to clouds, haze, or terrain [6]. Despite these difficulties, smoke is typically the first visible indicator of ignition, appearing well before flames are detectable, which makes robust early smoke detection a critical component of effective wildfire prevention and response [7].

With the rise of deep learning, vision-based object detectors became a central component of wildfire monitoring pipelines in both fixed-camera and aerial surveillance settings [8]. Convolutional neural network (CNN)-based architectures, including the R-CNN and YOLO families, were widely deployed due to their favorable accuracy-efficiency characteristics, while more recent transformer-based detectors and hybrid CNN-transformer models further advanced robustness and contextual modeling across diverse vision tasks [9–11]. Building on these developments, contemporary wildfire detection systems achieved substantial progress under a range of operational conditions [12]. However, detecting early-stage smoke remains a persistent challenge, as nascent plumes are often faint, low-contrast, and spatially small, with diffuse boundaries that can closely resemble confounding visual phenomena such as clouds, haze, or terrain textures [13]. A practical consideration in many deployments is that high-resolution frames are resized to fixed input dimensions to satisfy architectural constraints and enable efficient batched inference, which can attenuate subtle smoke cues and reduce their salience in downstream feature representations [14]. To mitigate these effects, a broad set of resolution-aware and small-object detection strategies has been investigated, including multi-scale inference, high-resolution processing, and image slicing/tiling, and these approaches demonstrated measurable gains for small targets in related remote-sensing and surveillance contexts [15]. Nonetheless, because real-world wildfire monitoring systems must often operate under strict latency and resource constraints, further improvements remain desirable—particularly in the early-smoke regime—toward enhancing recall and stabilizing confidence behavior near deployment thresholds while maintaining computational practicality [12].

To address the limitations of existing wildfire detection systems in the early-smoke regime, we propose a cost-efficient composite detection framework that improves sensitivity to small, faint smoke while preserving single-pass inference. The approach reallocates spatial resolution by embedding a targeted high-resolution region within a standard detector input, guided by a skyline-based spatial prior to emphasize likely smoke regions while retaining global scene context. For deployment, we incorporate a lightweight confidence refinement step that operates only in a low-confidence range to improve prediction stability without changing training or standard evaluation. The proposed pipeline avoids expensive multi-stage slicing or multi-scale inference while remaining compatible with real-time, resource-constrained settings. Our main contributions are:

- A composite multi-resolution representation that embeds high-resolution detail within a standard detector input using a single forward pass.
- A skyline-guided dynamic cropping strategy with reversible coordinate mapping for accurate localization in the original image space.
- A deployment-time confidence refinement mechanism to stabilize low-confidence early-smoke predictions.
- A comprehensive evaluation on early-smoke benchmarks (Early Smoke and Pyro-SDIS) demonstrating consistent gains in small-object and low-confidence regimes.

2. Related Work

2.1. Wildfire Monitoring Context

Wildfires arise from a combination of ignition sources and conducive environmental conditions, including low fuel moisture, high winds, and complex topography. Ignitions may be natural (e.g., lightning) or human-caused (e.g., equipment use, power infrastructure, and debris burning), and even small ignitions can rapidly evolve into large incidents under adverse weather and fuel conditions [16]. Once established, wildfires can spread quickly and produce severe societal and ecological impacts,

including loss of life and property, disruption to critical infrastructure, long-term ecosystem damage, and widespread degradation of air quality due to smoke transport [17]. These risks motivate an operational emphasis on early warning, where detecting nascent smoke plumes can enable faster dispatch and containment before initial ignitions escalate. Operational wildfire management relies on a combination of prevention, preparedness, and rapid response, supported by monitoring and situational awareness across multiple sensing modalities. Existing monitoring approaches are broadly categorized into human-based and autonomous systems. Human-based monitoring includes lookout towers, patrol flights, and public or field personnel reporting, which remain valuable but are labor-intensive, costly to scale, and inherently limited in temporal continuity and spatial coverage [18]. These limitations motivate increasing interest in automated alternatives that can operate persistently.

Autonomous monitoring systems span a range of sensing platforms operating at different spatial and temporal resolutions. Satellite-based systems provide wide-area coverage and routine global observation, making them effective for regional situational awareness but often limited in spatial resolution and revisit frequency for early-stage detection [19]. Aerial platforms, including crewed aircraft and unmanned aerial vehicles (UAVs), provide flexible, high-resolution sensing and rapid deployment capabilities, while their operational characteristics naturally introduce considerations related to cost, endurance, and weather dependence [20]. Ground-based vision systems, such as fixed CCTV camera networks deployed on towers or mountain ridges, enable continuous local surveillance and real-time visual confirmation, making them well suited for persistent monitoring of high-risk regions, while their effective coverage is influenced by terrain and camera placement [21]. Across these autonomous platforms, a wide range of analytical tools was explored. Early rule-based approaches relied on handcrafted color and motion heuristics to detect fire or smoke in video sequences [22]. Subsequent statistical learning approaches replaced fixed rules with engineered features and trained classifiers; for example, wavelet-based smoke characterization combined with support vector machines demonstrated reliable early detection with low false-alarm rates across surveillance scenarios [23]. More recently, AI-driven surveillance systems based on machine learning and deep learning gained growing attention, enabling end-to-end automated detection, reducing human workload and improving the timeliness and robustness of wildfire monitoring across heterogeneous sensing modalities [24].

2.2. Deep Learning for Wildfire Monitoring

Recent deep learning-based wildfire monitoring research was organized by the primary vision task and system structure employed [25]. A first line of work formulated wildfire monitoring as an image-level classification problem, predicting whether a frame contains fire or smoke to enable rapid triage under limited computation. For example, Akagic et al. [26] proposed LW-FIRE, a lightweight CNN designed for real-time wildfire image classification with dataset transformation to improve generalization. Seydi et al. [27] introduced Fire-Net, a deep learning-based remote sensing framework that integrates optical and thermal information to enable accurate and robust detection of active forest fires across diverse environments, including small-scale fire events.

A second line focused on object detection, aiming to localize fire or smoke regions with bounding boxes for spatially explicit situational awareness and downstream tracking [28]. Modern vision systems for wildfire monitoring commonly built on established object-detection families such as R-CNN and YOLO [9–11], which offered a practical balance between accuracy and real-time speed, alongside newer transformer-based and hybrid CNN-transformer designs that could better leverage broader scene context. Li et al. [29] proposed LEF-YOLO, a lightweight detector tailored for real-time identification of extreme wildfire cases and shown to perform effectively on a specialized dataset while maintaining fast inference. Bhargav and Singh [30] proposed a UAV-based wildfire detection framework that combines a CNN-based image classification model with YOLOv8 for object localization, demonstrating improved accuracy and computational efficiency over traditional methods while enabling timely fire detection from UAV imagery under challenging visual conditions.

A third line adopted semantic segmentation, producing pixel-level masks that could better represent irregular boundaries and diffuse phenomena, particularly for smoke plumes or elongated

flame fronts. Jonnalagadda and Hashim [31] developed a segmentation-based UAV pipeline optimized for real-time flame detection. Khryashchev and Larionov [32] developed a satellite-image wildfire segmentation model based on U-Net with a ResNet34 encoder, trained with augmentation on high-resolution RGB imagery and evaluated using Dice and IoU metrics to support early wildland fire detection.

Beyond single-task formulations, many practical systems used cascaded or multi-stage networks [33] that compose these tasks—most commonly using a lightweight classifier or coarse detector to propose candidate regions followed by a higher-capacity detector or segmenter to refine localization—improving robustness while controlling inference cost across high-resolution imagery. Ghali et al. [34] proposed a UAV-based wildfire framework that combines a deep ensemble classifier built on EfficientNet-B5 and DenseNet-201 with segmentation architectures including TransUNet, TransFire, and EfficientSeg. The framework enables accurate wildfire classification and precise pixel-level delineation, remaining robust to small fire regions and complex aerial backgrounds.

Collectively, these approaches have substantially improved the reliability of fire and smoke recognition in operational imagery; however, there remains significant scope for further improvement. In particular, a large portion of prior work is evaluated on scenarios in which the target signal is visually salient (e.g., large flames, dense smoke plumes, or fully developed fire fronts), where recognition cues are strong and localization is comparatively less ambiguous. Consequently, reported gains in classification accuracy, detection mAP, or segmentation overlap may not fully characterize performance in settings that are most critical for proactive monitoring, where the evidence is weaker and more readily confounded by background structure and atmospheric phenomena. Moreover, common efficiency-driven preprocessing practices—most notably resizing high-resolution frames to fixed input resolutions—may suppress fine-grained spatial detail that is informative for subtle visual signatures, motivating continued research on methods that generalize beyond prominent fire imagery while remaining computationally practical for real-time deployment.

2.3. Small Object Detection and Resolution Manipulation

Broadly, early wildfire detection shares key similarities with small object detection because the first visible cues—nascent smoke wisps or small flames—often occupy very few pixels and exhibit weak visual evidence. These cues are further degraded by downsampling and compression and are often confused with clouds, fog, dust, sun glare, or terrain textures, making them difficult to detect reliably in cluttered or low-contrast scenes. Remote-sensing surveys such as Cheng et al. [35] highlighted that even strong multi-scale and context-aware designs could struggle when targets became extremely small, which aligns closely with the "tiny smoke" regime encountered in early warning settings.

Consequently, detection systems for small objects tended to intervene on scale in two complementary ways. First, some works strengthened the detector itself to be more responsive to small targets, often through improved multi-scale handling and better localization behavior in the tiny-object regime [36–39]. Second, many practical pipelines manipulated the input or processing strategy so that early smoke or flame signals remained visible to the model without requiring uniformly higher resolution over the entire frame. Tiling and slicing approaches [40,41] and their refinements [42–46] reflected a common operational insight: allocating more spatial detail to candidate regions could materially improve detection when the event occupied only a small portion of the scene. Related region-adaptive methods similarly prioritized resolution where it mattered most by focusing computation on predicted areas of interest rather than treating all pixels equally [47].

Collectively, these directions improved small-target performance in fire monitoring, but they also exposed trade-offs that matter for early warning. Resolution-based pipelines may introduce additional runtime and engineering complexity due to repeated per-region inference and merging, and their performance may depend on design choices such as tile size and overlap. Meanwhile, detector-only improvements may raise sensitivity broadly across the full frame rather than explicitly preserving fine detail only in likely smoke regions. These issues are especially pronounced for early smoke, which

is diffuse and low-contrast and may occupy only a handful of pixels, motivating wildfire-specific strategies that allocate resolution and computation selectively while maintaining real-time practicality.

3. Methods

3.1. Overall Architecture

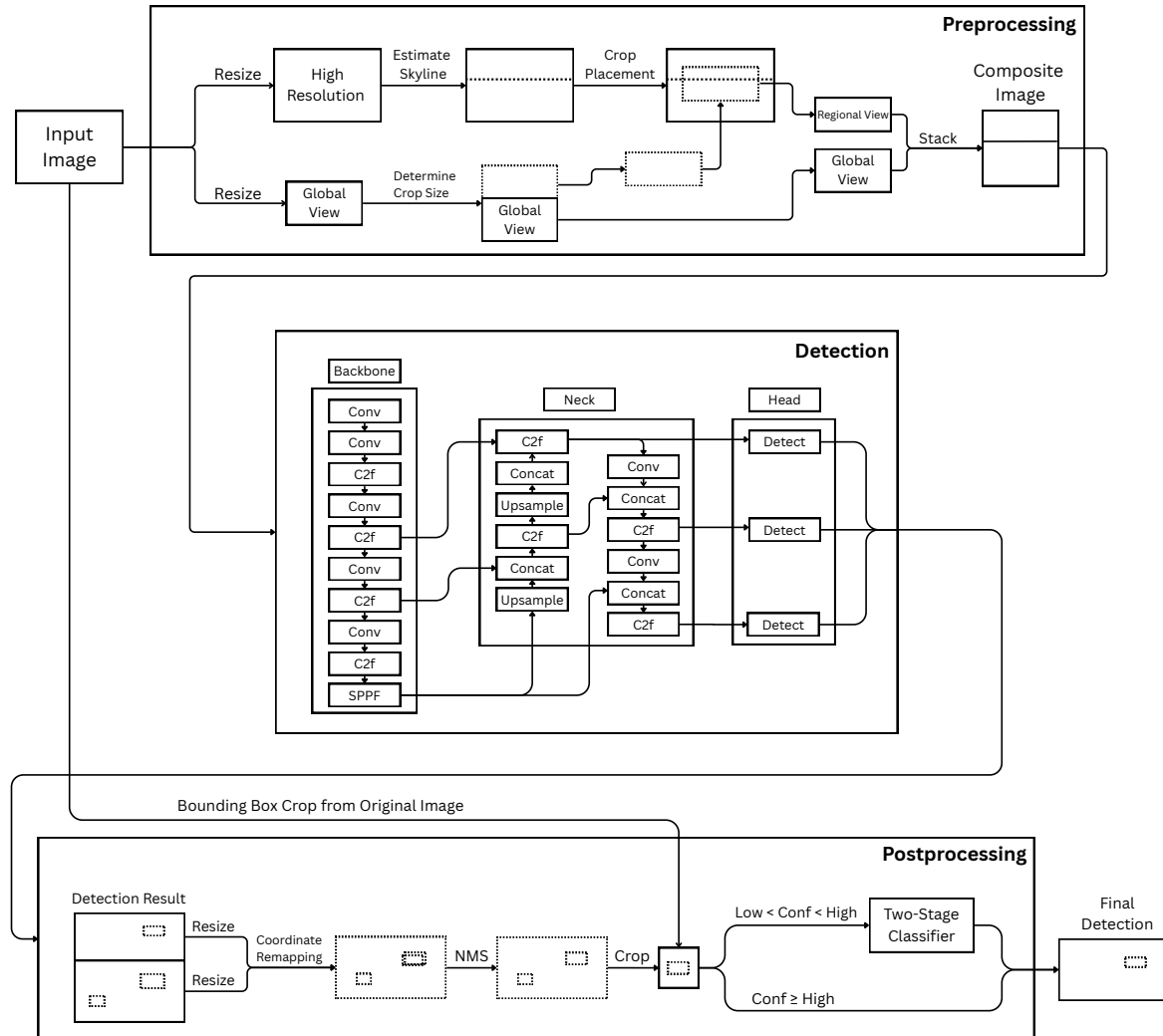


Figure 1. Overview of the proposed multi-resolution detection framework.

Figure 1 provides an overview of the proposed multi-resolution detection framework. The overall system is composed of three main components: (1) composite image generation with a sky-adaptive logic module for skyline-based cropping, (2) object detection on the composite input, and (3) a fusion stage that performs coordinate remapping, non-maximum suppression (NMS), and optionally a two-stage classifier for deployment. The following subsections describe each component in detail.

3.2. Composite Image Generation

Given an input image I , most modern object detectors operate on a fixed-size square input of $W_{in} \times H_{in}$ to enable efficient batching and to maintain a consistent feature-map geometry throughout the network. When the aspect ratio of I differs from the required square shape, the common practice is to resize I while preserving its aspect ratio and then apply letterboxing, i.e., padding the remaining area (typically with constant-valued gray or black pixels) so that the final tensor matches the $W_{in} \times H_{in}$ input constraint. These padded regions contribute negligible semantic information and primarily serve as a shape-normalization mechanism since they contain no scene content.

In this study, the otherwise uninformative padding area is repurposed as additional representational capacity. Instead of leaving this region as padding, the square canvas is filled with an auxiliary high-detail view of the same scene. This design preserves a global-context image while allocating more pixels to regions expected to contain early smoke signatures. The approach increases the effective spatial resolution available for small and visually ambiguous targets without changing the detector input size, modifying the backbone architecture, or incurring the computational overhead associated with multi-pass tiling or slicing-based inference (see Figure 2).

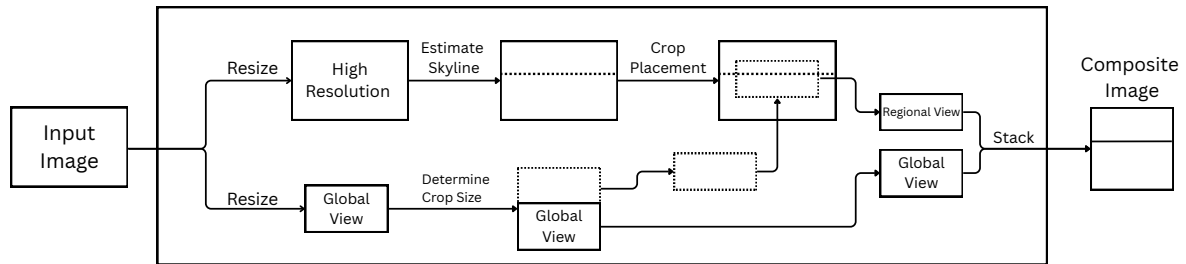


Figure 2. Composite image generation pipeline diagram.

To construct the composite representation, the original image I is first resized to a width of W_{in} pixels using bilinear interpolation while preserving aspect ratio, producing an image I_{global} of size $W_{in} \times H_{global}$. A second resized version I_{inter} is generated from the same high-resolution source using bilinear interpolation at a larger intermediate width S_{int} , allowing the system to retain object-level details at a higher spatial resolution. These two resized images provide the global and high-resolution views used in the composite. The cropping window is set to match the unused vertical space after resizing, i.e., a window of size $W_{in} \times (H_{in} - H_{global})$, so that stacking it with I_{global} yields a final $W_{in} \times H_{in}$ square. The vertical placement is determined using the estimated skyline position S , ensuring that the crop aligns with the region where smoke typically appears, and the crop is horizontally centered so that its midpoint aligns with the center of I_{inter} . Once the crop I_{roi} is extracted, it is placed above I_{global} to form the final composite image I_{comp} . The result is a $W_{in} \times H_{in}$ image composed of a top band containing a high-resolution regional view derived from I_{inter} and a bottom band containing the global view.

The overall procedure can be summarized as:

Algorithm 1 Composite Image Generation Procedure.

```

Procedure GenerateComposite( $I, W_{in}, H_{in}, S_{int}$ ) :
Input:  $I$  (RGB image),  $W_{in}, H_{in}$  (detector input size),  $S_{int}$  (intermediate resize width).
Output:  $I_{comp} \in \mathbb{R}^{W_{in} \times H_{in} \times 3}$  (composite for detector inference),
         $M$  (metadata for coordinate restoration).

1. Global view (base canvas for detector input):
    $I_{global} \leftarrow \text{ResizeWidth}(I, W_{in})$  (preserve aspect ratio),
    $I_{global} \in \mathbb{R}^{W_{in} \times H_{global} \times 3}$ .

2. Intermediate view (used to select a smoke-prior ROI):
    $I_{inter} \leftarrow \text{ResizeWidth}(I, S_{int})$  (preserve aspect ratio),
    $I_{inter} \in \mathbb{R}^{S_{int} \times H_{inter} \times 3}$ .

3. Define the ROI size to occupy the unused letterbox region:
    $H_{gap} \leftarrow H_{in} - H_{global}$ ,
    $(w_{roi}, h_{roi}) \leftarrow (W_{in}, H_{gap})$ .

4. Compute ROI location (skyline-guided, with fallback):
    $S \leftarrow \text{SkylineRow}(I_{inter})$ .
   if  $S$  is undefined:  $S \leftarrow \lfloor \eta H_{inter} \rfloor$  ( $\eta \in (0, 1)$  is a default prior).
    $m_0 \leftarrow \lfloor \frac{S_{int} - w_{roi}}{2} \rfloor$ ,  $m_1 \leftarrow m_0 + w_{roi}$ .
    $n_0 \leftarrow \text{clip}(\lfloor S - \beta h_{roi} \rfloor, 0, H_{inter} - h_{roi})$ ,
    $n_1 \leftarrow n_0 + h_{roi}$ .
   Define  $\text{ROI}_{inter} \leftarrow [m_0, m_1) \times [n_0, n_1)$ .

5. Crop the intermediate view:
    $I_{roi} \leftarrow \text{Crop}(I_{inter}, \text{ROI}_{inter})$ ,
    $I_{roi} \in \mathbb{R}^{W_{in} \times H_{gap} \times 3}$ .

6. Construct the detector input (single  $W_{in} \times H_{in}$  tensor):
    $I_{comp} \leftarrow \text{Concat}_{\text{vertical}}(I_{roi}, I_{global})$ ,
    $I_{comp} \in \mathbb{R}^{W_{in} \times H_{in} \times 3}$ .

7. Return composite and restoration metadata:
    $M \leftarrow \{W_{in}, H_{in}, S_{int}, H_{global}, H_{gap}, (m_0, n_0), \text{resize scales for } I_{global} \text{ and } I_{inter}\}$ .
   return  $(I_{comp}, M)$ .

```

During composite construction, a set of metadata parameters required for mapping detection boxes from composite coordinates back to the original image space is stored. These quantities capture the intermediate resizing scales, cropping boundaries, and any padding introduced during stacking. Let

- H_{gap} denotes the crop height,
- (m_0, n_0) the horizontal and vertical crop offsets,
- s_{int} and $s_{W_{in}}$ the scaling factors used for generating the intermediate-resolution and W_{in} -width images,
- (S_{int}, H_{inter}) the dimensions of the intermediate resized image,
- o_{top} and o_{bot} the vertical offsets introduced by cropping when mapping detections back to the original coordinate space.

These metadata variables are passed to the fusion stage, where they enable precise restoration of bounding boxes to the coordinate system of the original image I .

3.3. Dynamic Cropping

To obtain a stable reference point for placing the high-resolution crop, the skyline is estimated by leveraging color statistics in the YCrCb color space. The purpose of this module is to avoid placing the

crop window without guidance and instead use a consistent structural feature of outdoor scenes. The sky generally occupies the upper portion of the image and exhibits distinctive luminance–chrominance values; therefore, the transition between sky and terrain serves as a practical reference for positioning the higher-resolution crop. Figure 3 illustrates the intermediate and final outputs of this procedure.

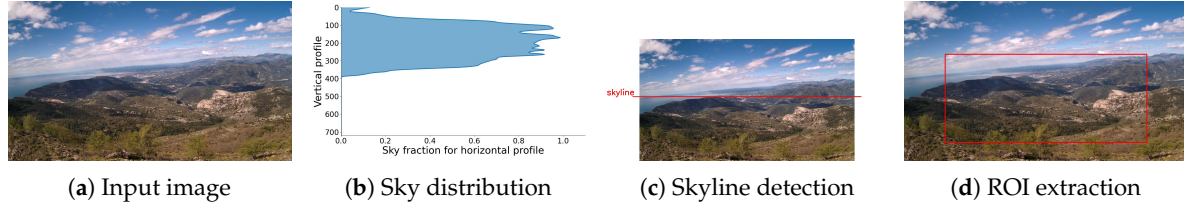


Figure 3. Step-by-step visualization of skyline detection. The module converts the image to YCrCb, identifies sky pixels via chrominance–luminance thresholds, and detects the sky-to-land transition through row-wise analysis.

To identify the skyline region, the high-resolution input image I_{inter} is first downsampled by a factor $\alpha \in (0, 1)$ to reduce computational cost, producing $I_{\text{inter},\alpha}$ whose width and height are scaled by α . The downsampled image $I_{\text{inter},\alpha}$ is then converted from RGB to YCrCb. In practice, we set $\alpha \in [0.125, 0.5]$ based on empirical validation, which substantially reduces computation while preserving accurate skyline estimation.

A binary mask is subsequently constructed by labeling pixels as “sky” according to luminance and chrominance criteria. Specifically, a pixel is marked as sky if its luminance exceeds the global mean of the Y channel and its chrominance values fall within empirically determined thresholds. The RGB-to-YCrCb conversion is calculated as:

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B, \\ Cb &= 128 - 0.168736R - 0.331264G + 0.5B, \\ Cr &= 128 + 0.5R - 0.418688G - 0.081312B, \end{aligned}$$

where $R, G, B \in [0, 255]$.

To determine the Cr and Cb thresholds, we assume $Y \in [0, 255]$ and partition the luminance range into three intervals using four breakpoints, $Y_1 = 64$, $Y_2 = 128$, $Y_3 = 160$, and $Y_4 = 192$. Sky regions are typically associated with higher luminance; therefore, pixels with $Y < Y_1$ (i.e., $Y < 64$) are discarded and are not considered as sky candidates. At each breakpoint Y_k , define empirically tuned chrominance thresholds

$$\Theta^{(k)} = \left(\tau_{Cb}^{\min,(k)}, \tau_{Cb}^{\max,(k)}, \tau_{Cr}^{\min,(k)}, \tau_{Cr}^{\max,(k)} \right), \quad k = 1, 2, 3, 4.$$

Each breakpoint threshold set $\Theta^{(k)}$ is initialized from dataset statistics by taking the sample mean and standard deviation of the sky-pixel chrominance distributions at luminance level Y_k , and setting each bound to lie within two standard deviations of the mean (i.e., $\mu \pm 2\sigma$). The final values are then selected empirically based on validation performance. For a pixel luminance $Y_{m,n}$, let $k \in \{1, 2, 3\}$ satisfy $Y_{m,n} \in [Y_k, Y_{k+1}]$, and define

$$\alpha_k(Y_{m,n}) = \frac{Y_{m,n} - Y_k}{Y_{k+1} - Y_k} \in [0, 1].$$

Then the luminance-adaptive thresholds are given by piecewise linear interpolation:

$$\begin{aligned}\tau_{Cb}^{\min}(Y_{m,n}) &= (1 - \alpha_k) \tau_{Cb}^{\min,(k)} + \alpha_k \tau_{Cb}^{\min,(k+1)}, \\ \tau_{Cb}^{\max}(Y_{m,n}) &= (1 - \alpha_k) \tau_{Cb}^{\max,(k)} + \alpha_k \tau_{Cb}^{\max,(k+1)}, \\ \tau_{Cr}^{\min}(Y_{m,n}) &= (1 - \alpha_k) \tau_{Cr}^{\min,(k)} + \alpha_k \tau_{Cr}^{\min,(k+1)}, \\ \tau_{Cr}^{\max}(Y_{m,n}) &= (1 - \alpha_k) \tau_{Cr}^{\max,(k)} + \alpha_k \tau_{Cr}^{\max,(k+1)}.\end{aligned}$$

The binary sky mask M is defined as:

$$M_{m,n} = \mathbb{I}[Y_{m,n} > \mu_Y \wedge \tau_{Cr}^{\min}(Y_{m,n}) \leq Cr_{m,n} \leq \tau_{Cr}^{\max}(Y_{m,n}) \\ \wedge \tau_{Cb}^{\min}(Y_{m,n}) \leq Cb_{m,n} \leq \tau_{Cb}^{\max}(Y_{m,n})],$$

where μ_Y denotes the average luminance across the image.

The row-wise profile is formalized by letting $r \in \{0, 1, \dots, H_{\text{inter},\alpha} - 1\}$ denote the image row index (measured from the top). Let $n_{\text{sky}}(r)$ be the number of pixels labeled as sky in row r . The row-to-row change in sky-pixel counts is then computed as $\Delta n_{\text{sky}}(r) = n_{\text{sky}}(r+1) - n_{\text{sky}}(r)$. This row-wise sky-pixel profile $n_{\text{sky}}(r)$, together with its discrete difference $\Delta n_{\text{sky}}(r)$, forms the basis for detecting abrupt transitions used to estimate the skyline location.

To verify that this transition is meaningful, we compare the total sky-pixel mass above and below a candidate skyline row S using the row-wise sky counts $n_{\text{sky}}(r)$. Specifically, we compute $S_{\text{above}} = \sum_{r < S} n_{\text{sky}}(r)$ and $S_{\text{below}} = \sum_{r > S} n_{\text{sky}}(r)$. If $S_{\text{above}}/S_{\text{below}} > \text{Th}_{\text{sky}}$ for a threshold Th_{sky} , the candidate S is accepted as the skyline; otherwise, no skyline is returned. The estimated skyline position is then used to guide the vertical placement of the cropping window in the composite generation process.

Given a detected sky boundary at S , the crop window with height H_{gap} is placed such that $n_{\text{top}} = S - H_{\text{gap}}/4$, ensuring that only a small portion of the crop window contains sky and the majority covers land or forest regions where smoke is more likely to originate. If no valid skyline is detected, the crop window is instead centered vertically within the image to provide a neutral fallback that avoids systematic bias.

3.4. Fusion: Remapping and Non-Maximum Suppression

Each detection need to be mapped back to the coordinate system of the original image, as inference is performed on the $W_{in} \times H_{in}$ composite image. This remapping reverses the sequence of resizing, cropping, and padding operations applied during composite construction. For every predicted bounding box, the appropriate transformation branch is selected based on whether the detection originates from the top band (high-resolution crop) or the bottom band (global view). The metadata recorded during composite generation provides the offsets, crop coordinates, and scaling factors required for restoration.

$(\hat{m}_1, \hat{n}_1, \hat{m}_2, \hat{n}_2)$ is used to denote the bounding box coordinates in composite-image pixels, with $\hat{m}_1$ and $\hat{m}_2$ corresponding to the left and right column indices and $\hat{n}_1$ and $\hat{n}_2$ corresponding to the top and bottom row indices. The corresponding coordinates in the original image are computed by $m_k^{\text{orig}} = \frac{\hat{m}_k - o_m + c_m}{s_m}$, $n_k^{\text{orig}} = \frac{\hat{n}_k - o_n + c_n}{s_n}$, with $k \in \{1, 2\}$. where (o_m, o_n) represent the padding offsets applied to the composite band, (c_m, c_n) represent the crop offsets in the intermediate-resolution image, and s_m, s_n denote the scaling factors used to produce the resized inputs. For detections originating from the top band (the high-resolution crop), the remapping parameters are $o_m = o_{\text{top}}$, $o_n = 0$, $c_m = m_0$, $c_n = n_0$, and $s_m = s_n = s_{\text{int}}$. For detections originating from the bottom band (the globally resized image), the corresponding parameters are $o_m = o_{\text{bot}}$, $o_n = H_{\text{gap}}$, $c_m = 0$, $c_n = 0$, and $s_m = s_n = s_{W_{in}}$. This ensures that all geometric transformations applied during composite construction are correctly reversed, allowing detections from both the high-resolution and global views to be restored to their precise locations in the coordinate frame of the original image I .

After remapping detections from both the top and bottom bands to the original image coordinates, duplicate bounding boxes may occur because the composite pipeline effectively simulates two views of the same scene. To consolidate these detections, standard non-maximum suppression (NMS) is applied with an IoU threshold τ_{NMS} . Let b_m denote the m -th bounding box, s_m its confidence score, and $\text{IoU}(b_m, b_n)$ the intersection-over-union between boxes b_m and b_n . Let N denote the total number of boxes. The IoU is defined as

$$\text{IoU}(b_m, b_n) = \frac{|b_m \cap b_n|}{|b_m \cup b_n|}.$$

Given an IoU threshold $\tau \in (0, 1)$, the set of selected boxes after NMS is

$$S = \{m \in \{1, \dots, N\} : \forall n \in \{1, \dots, N\}, s_n > s_m \Rightarrow \text{IoU}(b_m, b_n) < \tau_{\text{NMS}}\}.$$

The subset S forms the final set of detections for each image, expressed in the original coordinate frame.

3.5. Two-Stage Classifier

In deployment settings, a fixed confidence threshold is typically required to suppress false alarms. However, early-stage smoke plumes are typically faint, weakly textured, and visually diffuse, often occupying only a small portion of the image, which leads the detector to produce scores that cluster near the selected confidence threshold. Relying on a single cutoff therefore creates an inherent trade-off: raising the threshold improves precision but suppresses many true smoke detections, while lowering it increases recall at the cost of additional false positives. To alleviate this limitation, a lightweight two-stage classifier is introduced to refine only those detections whose confidence lies in an “uncertainty band” where YOLO predictions tend to be unreliable.

Detections with confidence below Th_{low} are treated as noise and discarded, whereas detections above Th_{high} are accepted without further processing. Only detections falling within the intermediate interval $[Th_{\text{low}}, Th_{\text{high}}]$ are passed to the classifier. This selective refinement follows the intuition of cascade-style detection frameworks, where ambiguous proposals receive additional evaluation, but differs in that our classifier is applied solely during deployment and does not modify detector training. For each detection requiring refinement, a square crop of size $r_{\text{cls}} \times r_{\text{cls}}$ is extracted from the original image such that the bounding-box center (m_c, n_c) lies at the center of the crop. Centering ensures that the classifier receives the local smoke structure along with the immediate surrounding context, while using a fixed resolution provides stable appearance statistics regardless of the underlying smoke size. The classifier then predicts whether the crop corresponds to foreground (smoke) or background.

The final decision rule combines the detector confidence with the classifier output. Detections with $\text{conf}_{\text{det}} < Th_{\text{low}}$ are discarded outright, and detections with $\text{conf}_{\text{det}} > Th_{\text{high}}$ are accepted immediately. For detections in the intermediate band, the classifier determines whether they are retained or removed. Computational overhead is minimized by limiting this refinement to a small subset of detections, so the additional computation is incurred only sporadically rather than uniformly across frames. Consequently, the pipeline maintains real-time throughput while improving robustness in cases where the detector alone yields insufficiently reliable predictions.

$$\text{Decide} = \begin{cases} \text{DISCARD}, & \text{conf}_{\text{det}} < Th_{\text{low}}, \\ \text{KEEP}, & \text{conf}_{\text{det}} > Th_{\text{high}}, \\ \text{KEEP}, & Th_{\text{low}} \leq \text{conf}_{\text{det}} \leq Th_{\text{high}} \text{ and } \text{cls}(d) = \text{fg}, \\ \text{DISCARD}, & \text{otherwise.} \end{cases}$$

4. Experimental Setup

4.1. Datasets

4.1.1. Dataset Composition

Publicly available wildfire datasets contain relatively few instances of early-stage smoke, and the majority of images depict large, fully developed fires that are not representative of early-detection conditions. Early-smoke scenes are inherently rare in real-world collections because ignition events are seldom captured precisely at onset, and smoke plumes remain visually faint for only a brief temporal window. In addition, many datasets include large numbers of near-duplicate frames, which can artificially inflate evaluation metrics and obscure true detector performance, particularly for small, low-contrast objects. Even when early-smoke images are present, the number of such samples is typically insufficient to support effective training. To address these limitations, a curated *Early Smoke* dataset is constructed from the FASDD-UAV and D-Fire collections. The Pyro-SDIS dataset is also employed, after deduplication, as a secondary benchmark to assess cross-dataset reliability.

4.1.2. Datasets Used

FASDD-UAV. FASDD is a large-scale flame and smoke detection dataset composed of three subsets (CV, UAV, RS) collected from terrestrial and airborne sensors such as surveillance cameras, watchtowers, and UAVs [48]. The UAV subset, FASDD_UAV, is used, which contains 25,097 UAV-captured images. Each image is annotated with bounding boxes in four formats (YOLO, VOC, COCO, TDML) for two object classes, *fire* and *smoke*, totaling 36,308 fire instances and 17,222 smoke instances. Image aspect ratios are primarily between 4:3 and 16:9. The dataset offers high-resolution aerial coverage but includes many large plumes and near-duplicate video frames; as a result, early-smoke examples are relatively scarce.

D-Fire. D-Fire is an image dataset for fire and smoke detection, comprising 21,527 images with 26,557 annotated bounding boxes, of which 11,865 are labeled as smoke and 14,692 as fire [49]. The classes are intentionally broad: smoke can appear concentrated or highly diffused, with color and intensity varying according to wind conditions and the materials being burned. Images were collected from the Internet, controlled fire simulations in the Technological Park of Belo Horizonte, surveillance cameras at the Universidade Federal de Minas Gerais (UFMG), and cameras at the Serra Verde State Park in Belo Horizonte. Additional synthetic images were created by compositing smoke patches onto green landscape backgrounds, a form of augmentation previously shown to be effective for training on synthetic smoke images. Annotations follow the YOLO format with normalized coordinates in $[0, 1]$, and the dataset spans diverse fire and smoke patterns, scenarios, camera viewpoints, and lighting conditions. As with FASDD_UAV, many scenes contain large flames; while some high-resolution early-smoke examples exist, most images depict well-developed fires, thus additional filtering is necessary for early-smoke-focused experiments.

Pyro-SDIS. Pyro-SDIS is a wildfire smoke detection dataset developed with the French Fire and Rescue Services (SDIS) and volunteers from the Pyronear association [50]. Images are captured from fixed detection towers equipped with 4–5 high-resolution cameras and an on-site microcomputer, monitoring mostly forested environments in wildfire-prone areas. Smoke is typically far from the camera, small in the image, and often low-contrast under diverse lighting (from moderately dark to bright) and weather conditions (sunny, cloudy, hazy), making detection particularly challenging. The current release contains 33,636 images in total, including 28,103 images with smoke and 31,975 smoke instances, annotated in YOLO format by Pyronear volunteers. Unlike FASDD_UAV and D-Fire, Pyro-SDIS primarily captures early-stage wildfire smoke, but the raw data include many near-duplicate frames and some annotation inconsistencies, so additional preprocessing is required before use in this study.

4.1.3. Dataset Construction

Two derived datasets are constructed for the experiments. The *Early Smoke* dataset merges FASDD_UAV and D-Fire, then applies near-duplicate removal and smoke-focused filtering, while a separately deduplicated *Pyro-SDIS* split is built from the raw Pyro-SDIS release without additional smoke filtering.

Near-duplicate frames are removed to reduce metric inflation using a perceptual hashing (pHash) procedure based on low-frequency DCT features. Each image I is converted to grayscale and resized to 32×32 pixels using bilinear interpolation, producing a normalized representation I' that suppresses high-frequency variation. A 2D Discrete Cosine Transform is applied to I' , and only the top-left 8×8 coefficients are retained because they capture the dominant low-frequency structure. Let $D = \{C(u, v) \mid 0 \leq u, v < 8\}$ denote this block. Then, the median value m of all coefficients in D is computed, and a 64-bit binary hash is generated by setting

$$h(u, v) = \begin{cases} 1, & C(u, v) > m, \\ 0, & C(u, v) \leq m, \end{cases}$$

followed by flattening the result in row-major order. Similarity between two images is evaluated via the Hamming distance between their hashes:

$$d(H_1, H_2) = \sum_{k=1}^{64} \mathbf{1}(H_1^k \neq H_2^k).$$

Images with $d(H_1, H_2) \leq 3$ are considered duplicates in the Early Smoke dataset, and $d(H_1, H_2) \leq 10$ flags duplicates within Pyro-SDIS. These empirically chosen thresholds, validated through extensive visual inspection, reliably remove near-identical frames—particularly those from consecutive UAV captures or repeated monitoring scenes—without discarding distinct examples.

After deduplication, additional filtering is applied to the Early Smoke dataset to isolate early-smoke instances. Pyro-SDIS is already curated for early-smoke detection and does not include the large-fire imagery commonly observed in FASDD-UAV and D-Fire; therefore, no additional smoke-specific filtering is applied to that split. This filtering procedure is intended to remove large-fire scenes while retaining images in which smoke is faint, spatially limited, and indicative of early ignition.

First, the annotated smoke region is required to occupy less than 2% of the total image area. Visual inspection of FASDD-UAV and D-Fire suggests that genuine early-smoke plumes typically occupy well under 1%; however, a 2% threshold is adopted to preserve coverage while excluding larger fire events. Second, each image is required to contain exactly one bounding box, consistent with the single ignition source expected in early-stage wildfires; images with multiple boxes are frequently associated with larger or fragmented fire scenes. Only images annotated with the smoke bounding-box label are retained, as visible flame is uncommon during the earliest ignition stages. The dataset is further restricted to high-resolution images with a longer side exceeding 1080 px, ensuring that multi-resolution composite generation can be performed without sacrificing spatial detail. These filtering steps yield a tightly curated Early Smoke set that reflects genuine early-smoke conditions and forms the foundation for evaluating fine-grained smoke detection.

4.1.4. Dataset Description and Statistics

The resulting datasets after the construction steps are summarized in Section 4.1. The filtered *Early Smoke* subset drawn from FASDD-UAV and D-Fire contains 936 images in which smoke is small ($< 2\%$ of image area), faint, and set against varied backgrounds. The deduplicated Pyro-SDIS subset contains 3,523 unique images captured by tower-mounted multi-camera systems in forested settings, where smoke typically appears distant and low-contrast under diverse lighting and weather conditions; all instances are annotated in YOLO format by Pyronear volunteers. Accordingly, two datasets are constructed:

1. **Early Smoke Dataset:** 936 carefully filtered early-smoke images drawn from FASDD-UAV and D-Fire, emphasizing small, faint, low-contrast plumes with $< 2\%$ image area under varying illumination and high background complexity.
2. **Pyro-SDIS Dataset:** 3,523 deduplicated images from tower-mounted camera systems in predominantly forested scenes, containing far-distant, low-contrast smoke under diverse lighting and weather conditions, with manually curated YOLO annotations by Pyronear volunteers.

The two datasets serve complementary roles in our evaluation. The Early Smoke Dataset provides clean, controlled early-smoke instances tailored for training and detailed analysis of model behavior on canonical early-stage plumes. The Pyro-SDIS subset introduces a meaningful distribution shift in terms of hardware, geography, appearance, and atmospheric conditions, making it well suited for testing robustness and generalization under realistic deployment scenarios. Qualitative examples for each dataset are illustrated in Figures 4 and 5, respectively.

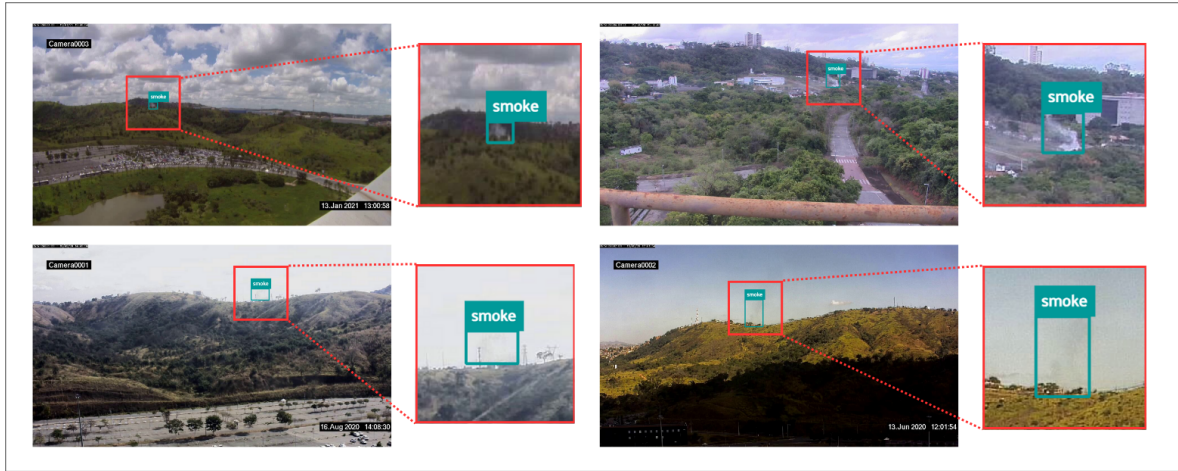


Figure 4. Qualitative examples from the Early Smoke dataset.

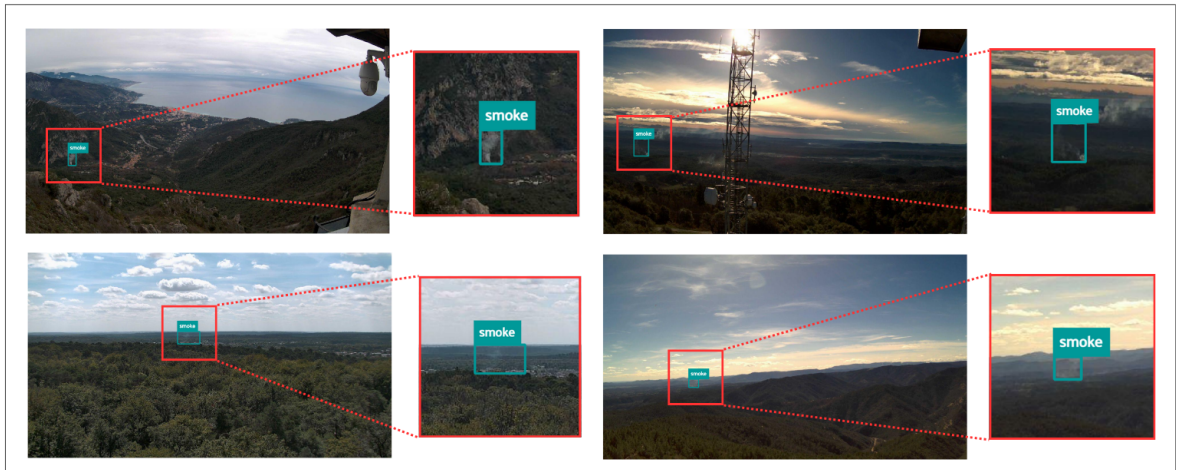


Figure 5. Qualitative examples from the Pyro-SDIS dataset.

4.2. Training for Early Smoke Detection

All experiments are conducted on a single workstation environment; Table 1 summarizes the hardware and software configuration used consistently across detector and classifier training to ensure reproducibility.

Table 1. Hardware and software environment for all experiments.

Component	Specification
GPU	NVIDIA RTX A5000 (24GB)
CPU & Memory	Intel(R) Core(TM) i5-14400F; 64GB
OS	Ubuntu 24.04.3 LTS
CUDA version	12.6
Python version	Python 3.10.18
PyTorch version	2.7.1+cu126

The training pipeline proceeds in two stages. First, the composite-based detector is trained on both datasets using identical train/validation/test splits and the same composite-generation procedure used at inference to align distributions. Second, a post-hoc binary classifier is trained on detector proposals cropped from the original-resolution images to reduce false positives while preserving recall on faint, early-stage smoke.

4.2.1. Detector Training

The approach is evaluated using two curated datasets that focus on small, early-smoke instances. The first is the Early Smoke Dataset, consisting of 936 images obtained by filtering FASDD-UAV and D-Fire as described in Section 4.1. The second is the deduplicated Pyro-SDIS dataset, containing 3,523 unique images.

All datasets are split at the image level into training, validation, and test sets using a 70/10/20 ratio. Test images are held out strictly for final evaluation and are never used for hyperparameter tuning; all model selection is performed using validation performance. Ground-truth annotations follow the standard YOLO format, consisting of class ID, box center, and box dimensions, all normalized to $[0, 1]$. During training and validation, each image is converted into a $W_{in} \times H_{in}$ composite using the procedure described in Section 3. The composite dimension is set to $W_{in} = H_{in} = 640$, selected empirically and chosen to match the widely used 640×640 detector input size in related object-detection pipelines. The ground-truth bounding box is used to determine the placement of the high-resolution crop, ensuring that the smoke region lies fully within the crop window. This design ensures that each annotated plume appears in both the global band and the high-resolution band of the composite, providing consistent supervision for detector training. Training is performed directly on composite images because inference in the deployed system is also conducted on composites. This alignment reduces distribution mismatch between training and deployment and avoids additional covariate shift introduced by composite generation. Test images remain at their original resolution and are converted to composite form only at inference time. To improve robustness to positional variation, three composite variants are generated for every training and validation image by shifting the crop horizontally so that the smoke region appears near the left ($\approx 25\%$ of width), center ($\approx 50\%$), or right ($\approx 75\%$). This augmentation reflects deployment conditions, where the sky-adaptive logic may place the crop at different horizontal locations depending on the skyline estimate. In addition to positional augmentation, the default Ultralytics YOLO augmentations are applied to the training split only.

For all experiments, a YOLOv8s detector is trained for 100 epochs with a batch size of 16 and an initial learning rate of 10^{-2} , initialized from pretrained weights. The Early Smoke Dataset is optimized using AdamW, whereas the Pyro-SDIS dataset is optimized using SGD, reflecting the configurations that yielded the most stable convergence in each setting. These hyperparameters follow standard YOLO training practice and support reliable optimization across both small, low-contrast early-smoke instances and far-distance smoke scenes present in Pyro-SDIS.

During inference, the YCrCb thresholds for skyline estimation (Section 4.1.4) are fixed to empirically tuned values that stabilize skyline detection on the training images. Inputs are assumed to be RGB images. The sky mask retains pixels whose Cb and Cr values fall within luminance-adaptive ranges, where the lower and upper thresholds are obtained by linear interpolation across four lumi-

nance breakpoints, $Y = \{64, 128, 160, 192\}$. These breakpoint ranges were determined empirically by inspecting the Cb/Cr distributions of pixels corresponding to sky regions in the images used in our datasets. Specifically, the breakpoint values are set to $Cb \in [115, 162]$ and $Cr \in [110, 132]$ at $Y = 64$, $Cb \in [129, 186]$ and $Cr \in [96, 124]$ at $Y = 128$, $Cb \in [121, 183]$ and $Cr \in [93, 130]$ at $Y = 160$, and $Cb \in [118, 168]$ and $Cr \in [96, 136]$ at $Y = 192$. For a pixel with luminance Y between adjacent breakpoints, each of $\tau_{Cb}^{\min}(Y)$, $\tau_{Cb}^{\max}(Y)$, $\tau_{Cr}^{\min}(Y)$, and $\tau_{Cr}^{\max}(Y)$ is computed via piecewise linear interpolation between the corresponding endpoint values, and the pixel is labeled as sky if $\tau_{Cb}^{\min}(Y) \leq Cb \leq \tau_{Cb}^{\max}(Y)$ and $\tau_{Cr}^{\min}(Y) \leq Cr \leq \tau_{Cr}^{\max}(Y)$. A visualization of these thresholds is shown in Figure 6. The acceptance-ratio threshold is set to $Th_{sky} = 5.0$, and the skyline-estimation downsampling factor is set to $\alpha = 0.25$. These settings remain constant across all reported experiments.

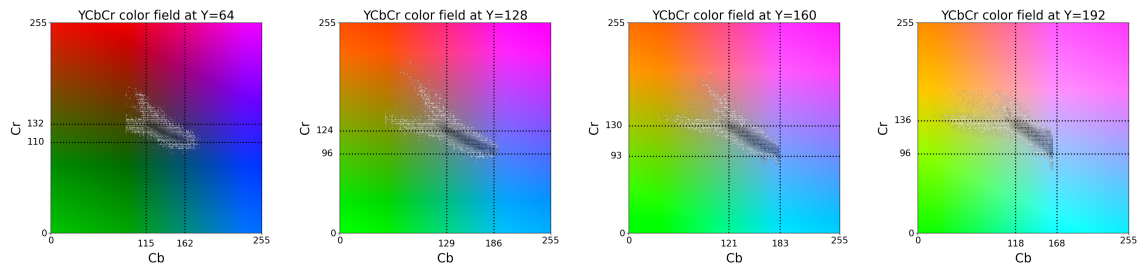


Figure 6. YCbCr planes at the four luminance breakpoints used to define the luminance-adaptive chrominance thresholds.

4.2.2. Training for Early Smoke Classification

To refine detector outputs and suppress low-confidence false positives, a binary classifier is trained on localized image crops derived from detector predictions. The classifier operates on 224×224 patches extracted from the original images and predicts whether each patch contains wildfire smoke.

The classifier is trained on the training splits of both the Early Smoke and Pyro-SDIS datasets. For each image, the same composite-generation pipeline used during detector inference is applied to maintain consistency between detector and classifier inputs. After the composite-based detector is trained, it is run on the training portions of both datasets and the resulting predicted bounding boxes are collected. These detector-generated proposals are then converted into 224×224 crops extracted from the original-resolution images and centered on each predicted region. By using detector proposals rather than ground-truth boxes, the classifier is trained on the same type of inputs encountered during deployment. This process yields a diverse set of localized patches containing both true smoke plumes and detector-induced false positives.

Each crop is assigned a label based on its intersection-over-union (IoU) with the ground-truth smoke annotations. Crops with $\text{IoU} > 0.5$ are labeled as foreground (true smoke), whereas crops with $\text{IoU} < 10^{-6}$ are labeled as background (false positives). Crops with intermediate IoU values are discarded to mitigate ambiguity in supervision. After labeling, all crops are randomly partitioned into a 70/10/20 train/validation/test split, with splitting performed after crop generation to prevent leakage across splits. All classifier crops are extracted directly from the original-resolution images rather than from the composite representation. Cropping windows are centered on the detector's predicted bounding boxes, with appropriate boundary handling applied when detections occur near image borders. The resulting crop dataset is class-imbalanced, as detector-generated proposals contain substantially more background patches than true smoke instances. To mitigate this imbalance, the foreground samples are augmented until the numbers of positive and negative crops are approximately matched. The augmentation pipeline applies horizontal flips with probability 0.5 and introduces mild geometric perturbations, including small translations, slight scale variations, and rotations up to 15° . These transformations increase sample diversity while preserving the visual characteristics of early-stage smoke. Augmentation is applied only to the training split to keep the validation and test distributions unchanged.

For all classifier experiments, a YOLOv8-s classification model is trained, initialized from the pretrained yolov8s-cls.pt weights for 100 epochs with a batch size of 32 and an initial learning rate of 10^{-2} . The model is optimized using the AdamW optimizer, which provided stable convergence on the balanced foreground–background crop dataset derived from detector proposals. These hyperparameters follow standard YOLO classification practice and are sufficient to train a robust classifier capable of discriminating true smoke plumes from detector false positives.

4.3. Evaluation Metrics

Detection performance is evaluated using standard object detection metrics. Precision measures the fraction of predicted detections that are correct, while recall measures the fraction of ground-truth objects that are successfully detected. Both are defined as below where TP, FP, and FN denote true positives, false positives, and false negatives, respectively. The F1 score, which summarizes precision and recall into a single harmonic-mean metric.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Average Precision (AP) quantifies detection performance by integrating the precision–recall curve at a specified IoU threshold. Experiments report AP at fixed thresholds (e.g., AP@0.5) as well as the COCO-style AP averaged over IoU thresholds from 0.50 to 0.95 in increments of 0.05. Mean Average Precision (mAP) is computed as the mean of AP values across the specified IoU thresholds and is used as the primary detection metric.

In addition, size-specific COCO metrics are also reported, including AP_{small} , $\text{AP}_{\text{medium}}$, and AP_{large} . These categories correspond to objects with areas smaller than 32^2 , between 32^2 and 96^2 , and larger than 96^2 pixels, respectively. These metrics provide insight into detection performance across different object scales.

5. Results and Discussion

For completeness, performance is evaluated on the Early Smoke and Pyro–SDIS datasets using (i) standard scientific metrics and (ii) a deployment setting with a fixed confidence threshold and uncertainty-band refinement. The baseline is a YOLOv8s detector trained on 640×640 letterboxed inputs without composite processing.

5.1. Overall Detection Performance

As shown in Table 2, our pipeline improves performance across all AP metrics, with the most notable gains in AP@0.5 and AP_{small} on the Early Smoke Dataset. The proposed method increases AP@0.5 from 0.818 to 0.852, resulting in a net improvement of 3.4%. The enhancement is most pronounced for small objects: AP_{small} increases from 0.054 to 0.212, representing nearly a fourfold improvement in the smallest-object regime. These improvements are more clearly reflected in the curves. As shown in Figure 7, the recall curves indicate relative gains of up to 60% in the high-confidence region (0.7–0.9). This behavior yields an overall increase of approximately 20% in average recall. These recall gains are accompanied by an expected reduction in precision of approximately 10% over the same confidence range. This trade-off is considered acceptable given the substantially larger recall improvement and the central importance of recall in early-smoke detection tasks. Additional evidence is provided by the ΔF1 curve, which indicates a net F1 increase of approximately 2%.

Table 2. Detection performance on the Early Smoke dataset.

Model	AP@0.5:0.95	AP@0.5	AP@0.75	AP Small	AP Medium	AP Large
Baseline	0.374	0.818	0.285	0.054	0.277	0.456
Ours	0.420	0.852	0.388	0.212	0.323	0.493

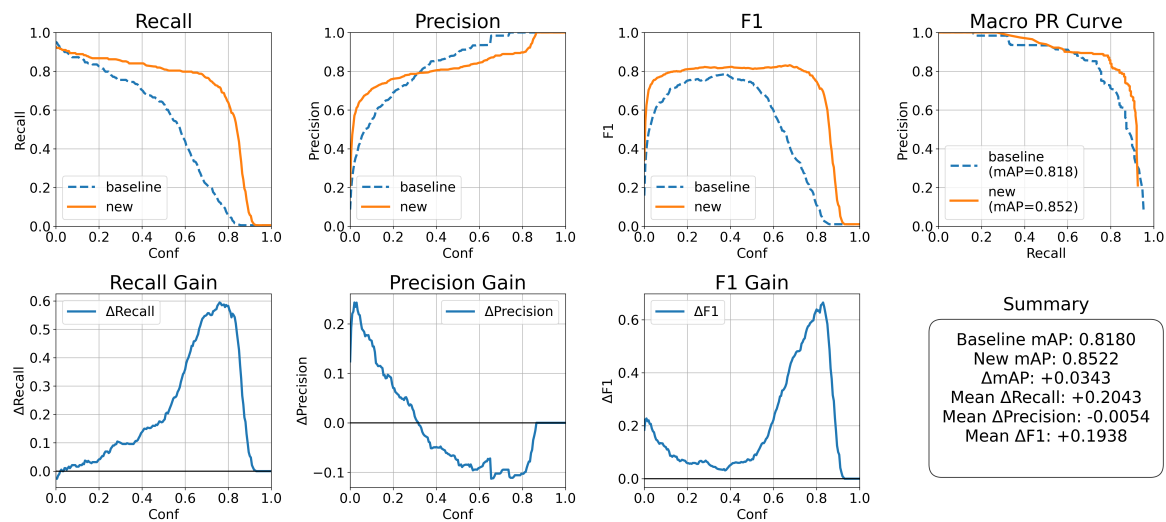


Figure 7. Comparison of baseline vs. composite detector performance on the Early Smoke dataset.

Beyond the baseline comparison, evaluations against other mainstream detector architectures indicate that the proposed pipeline consistently outperforms competing alternatives. As summarized in Table 3, the proposed method achieves the highest AP@0.5:0.95, AP@0.5, and AP@0.75. It also attains the best performance across all size-specific AP metrics. YOLOv5su exhibits strong AP@0.5 performance (0.839) but yields a comparatively lower AP_{small}. By contrast, YOLOv10s attains AP_{small} performance comparable to the proposed method (0.210) but underperforms on AP@0.5. The proposed method delivers balanced and superior results across metrics, supporting the effectiveness of composite-based training for early-stage wildfire smoke detection.

Table 3. Comparison against other mainstream detectors on the Early Smoke dataset.

Model	AP@0.5:0.95	AP@0.5	AP@0.75	AP Small	AP Medium	AP Large
YOLOv5su	0.384	0.839	0.277	0.136	0.304	0.453
YOLOv8s	0.374	0.818	0.285	0.054	0.277	0.456
YOLOv9s	0.321	0.776	0.172	0.077	0.221	0.400
YOLOv10s	0.367	0.804	0.315	0.210	0.296	0.437
YOLOv11s	0.379	0.792	0.319	0.134	0.265	0.465
Ours	0.420	0.852	0.388	0.212	0.323	0.493

To complement the quantitative results, Figures 8 and 9 present qualitative examples from the Early Smoke dataset. These examples illustrate the characteristic difficulty of the task. Smoke plumes are extremely small, visually faint, and frequently embedded in complex backgrounds. In many cases, baseline detectors either miss the plume entirely or produce low-confidence predictions. In contrast, the composite-trained model yields more consistent detections with higher confidence. This effect is most evident in scenes where the smoke occupies only a few dozen pixels. These examples provide visual evidence for the improvements reported in Table 2.

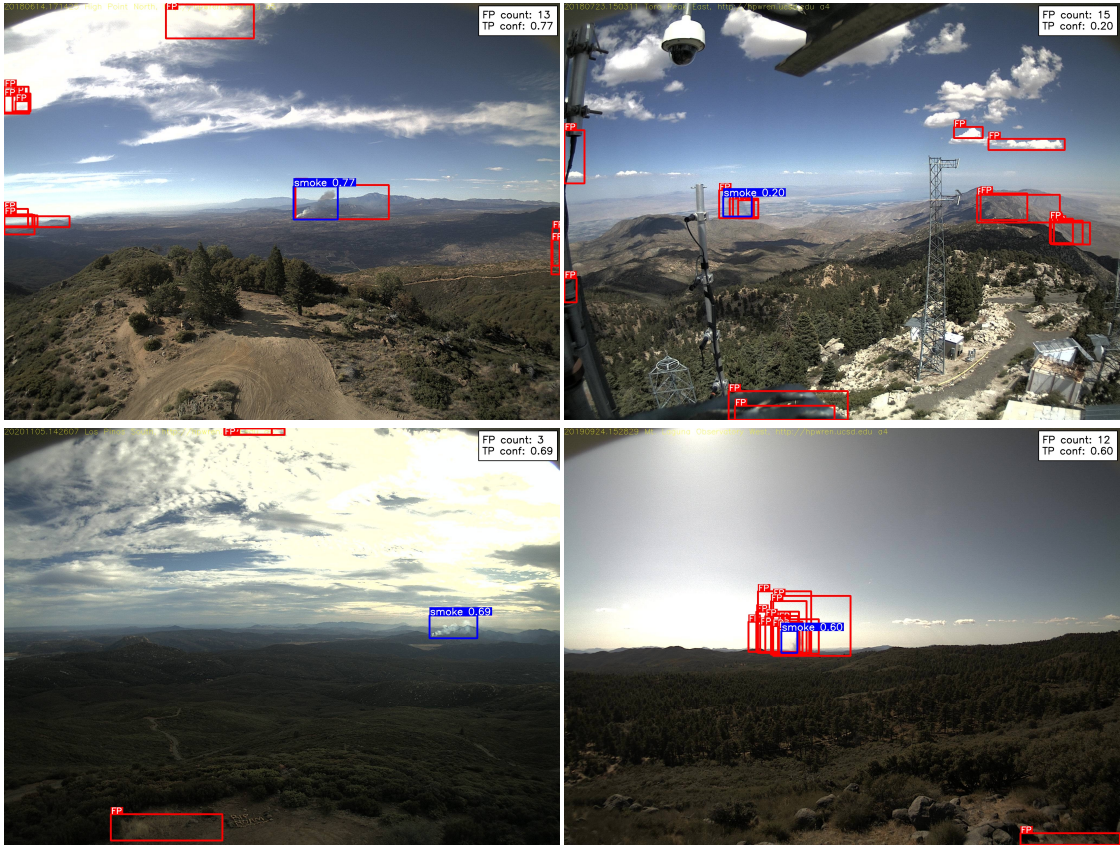


Figure 8. Qualitative detection examples on the Early Smoke dataset using baseline YOLOv8s detection. Bounding boxes are color-coded by detection outcome: blue denotes true positives and red denotes false positives.

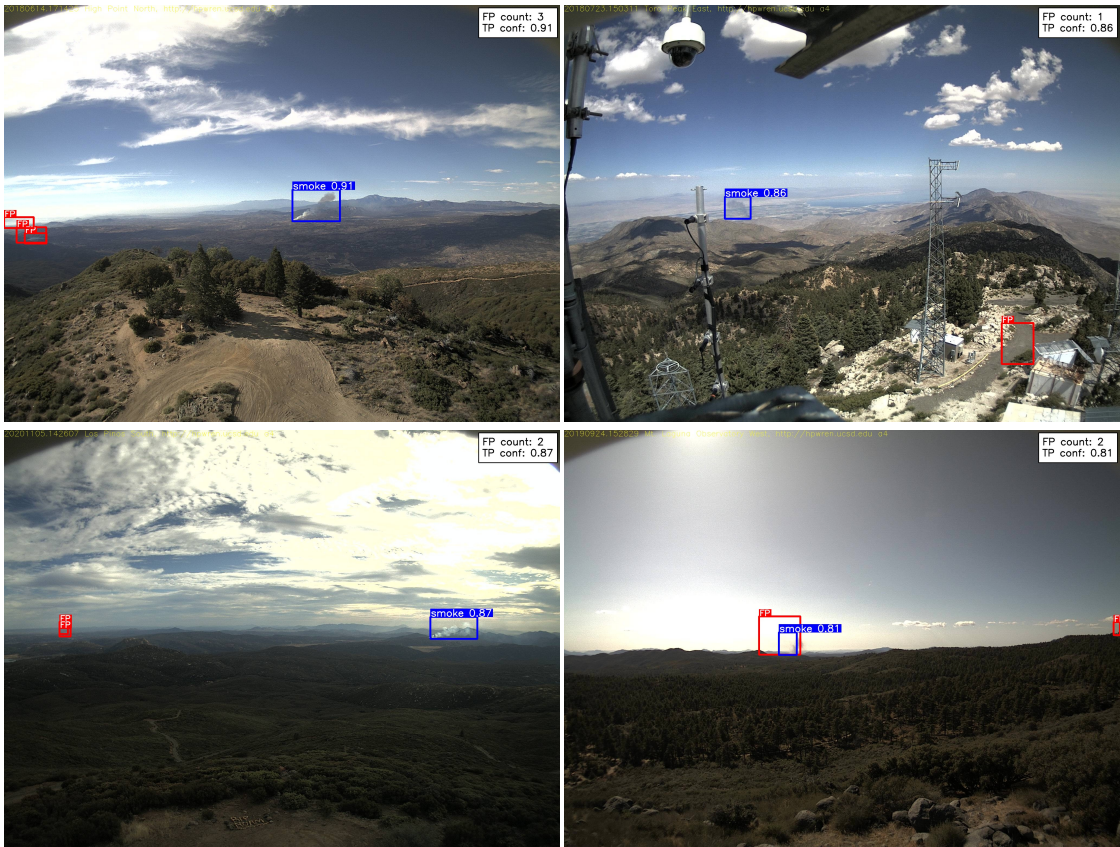


Figure 9. Qualitative detection examples on the Early Smoke dataset using the composite-trained model. Composite training improves confidence stability and plume localization for small smoke plumes.

Collectively, these results demonstrate that the proposed pipeline effectively enhances detection performance on small smoke objects that are critical for early wildfire detection. By leveraging multi-resolution compositing during training, the detector learns to better capture fine-grained smoke features while retaining global context, leading to substantial improvements in both recall and AP metrics, particularly for the smallest smoke plumes that are most challenging to detect. This indicates that the pipeline is especially effective under the difficult visual conditions of early-smoke detection, where standard single-resolution models struggle to preserve sufficient fine-scale structure.

5.2. Deployment-Only Evaluation (Two-Stage Classifier)

In deployment settings, the detector operates under a fixed confidence threshold that balances false alarms and missed detections. Early-smoke predictions often lie near the decision boundary, where confidence values fluctuate and standard thresholding becomes unstable. To address this issue, a two-stage classifier is introduced to refine only detections whose confidence falls within a predefined uncertainty band. This refinement is applied exclusively at deployment time and does not affect the detector’s scientific evaluation reported in the previous section.

Confidence Threshold Selection

Figure 10 shows the confidence–recall and confidence–precision curves computed for the Early Smoke dataset. The detector exhibits a sharp drop in recall only below a confidence of approximately 0.2, making $Th_{conf} = 0.2$ a suitable operating threshold for recall-oriented deployment. Precision remains relatively unstable across the low-confidence region due to the extremely small and visually faint nature of the plumes. Consequently, a large proportion of true detections fall within the 0.1–0.3 interval, reinforcing this band as the appropriate target for classifier refinement. The selected thresholds therefore balance recall preservation with minimal disruption to detector outputs while allowing the classifier to correct uncertain cases.

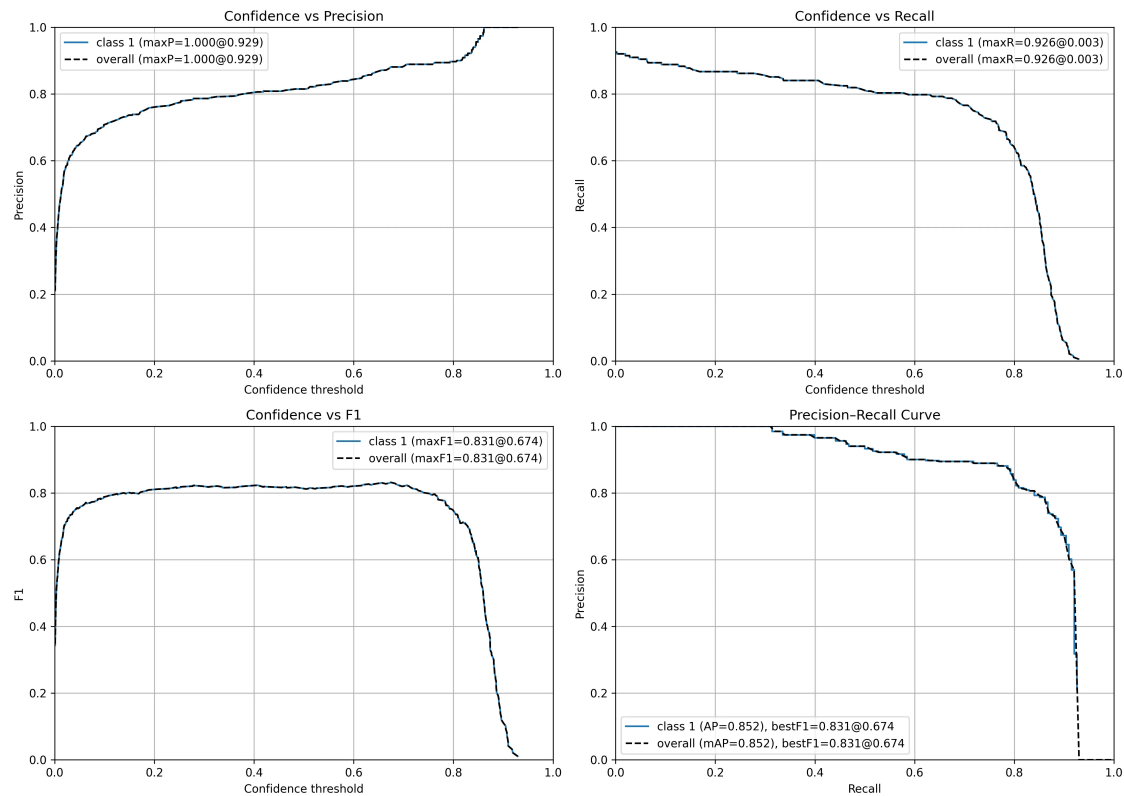


Figure 10. Confidence–recall and confidence–precision curves for the Early Smoke dataset. The selected deployment threshold of $Th_{conf} = 0.2$ preserves recall while the 0.1–0.3 band defines the refinement region for the two-stage classifier.

Table 4 reports deployment-setting performance on the Early Smoke dataset under a fixed confidence threshold of $Th_{conf} = 0.2$. The multiresolution detector substantially outperforms the baseline YOLOv8 across all metrics, including $AP_{50:95}$, AP_{75} , and AP_S , establishing a stronger operating point under fixed-threshold filtering. When the two-stage classifier is applied to detections within the 0.1–0.3 uncertainty band, additional gains are observed. These gains are driven by more reliable handling of borderline predictions, which are most sensitive to hard-threshold decisions.

Notably, this selective refinement recovers true positives that would otherwise be discarded when using a single detector score for thresholding. In operational settings, detections near Th_{conf} often include visually ambiguous but correct early-smoke instances; the classifier provides an additional verification signal that can promote these candidates above the decision threshold while suppressing low-confidence false positives. As a result, the two-stage system improves overall AP and especially strengthens small-object performance without requiring a lower global threshold, making it more closely aligned with deployment constraints where fixed-threshold decisions are necessary. Figures 11 and 12 present qualitative examples comparing baseline detections against the full pipeline with multiresolution detection and two-stage classification.

Table 4. Deployment evaluation on the Early Smoke dataset at a fixed threshold of $Th_{conf} = 0.2$. The two-stage classifier refines detections in the 0.1–0.3 confidence interval.

Experiment	$AP_{50:95}$	AP_{50}	AP_{75}	AP Small	AP Medium	AP Large
Multiresolution Only	0.404	0.813	0.378	0.188	0.313	0.475
Multiresolution + Classifier	0.410	0.820	0.383	0.188	0.318	0.482

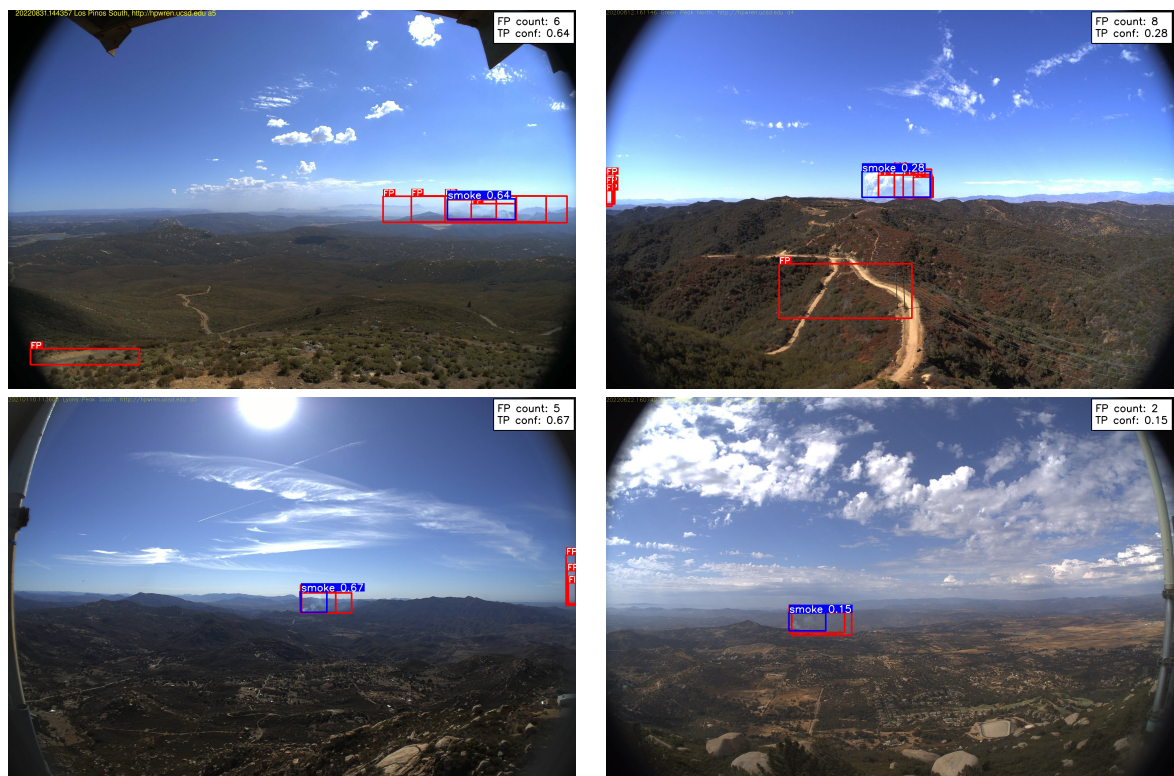


Figure 11. Qualitative ablation examples using the baseline detector. Bounding boxes are color-coded by outcome: blue denotes true positives and red denotes false positives.

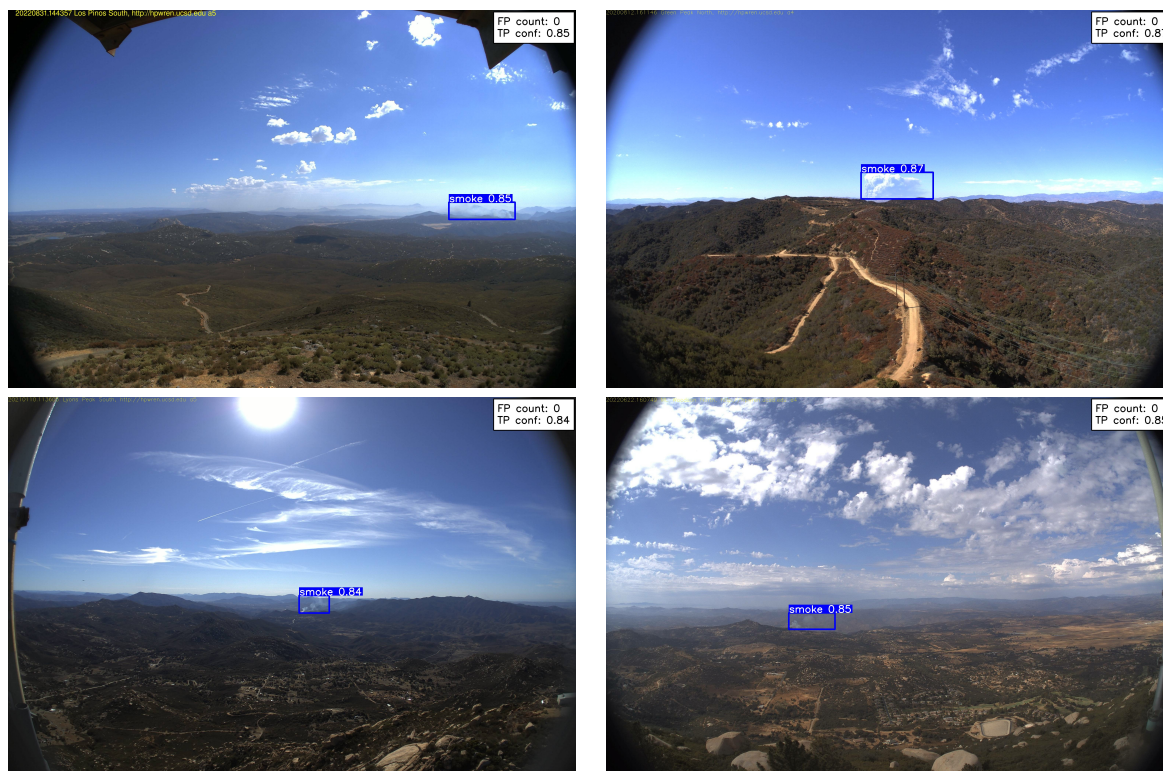


Figure 12. Qualitative ablation examples using the full pipeline.

5.3. Ablation Study

To quantify the contribution of each module and to characterize the accuracy–efficiency trade-off of the complete system, we conduct an ablation study in which the multiresolution detector and the two-stage classifier are enabled individually and in combination. Such analysis is required to attribute performance gains to specific design choices and to identify the primary sources of additional latency under full deployment. In particular, because the second-stage classifier introduces extra computation, the ablation results provide empirical justification for a selective, event-driven invocation strategy, in which the classifier is applied only to ambiguous detector outputs rather than uniformly across all frames, thereby limiting overhead during background-dominated surveillance. Table 5 reports an ablation study analyzing the individual and combined effects of the multiresolution detector and the two-stage classifier on the Early Smoke dataset. The baseline YOLOv8 model exhibits very low performance on small objects ($AP_S = 0.037$), reflecting the extreme difficulty of detecting early-stage smoke plumes that occupy only a few pixels and lack strong visibility.

Introducing the two-stage classifier alone yields only marginal changes (slight gains in AP_{50} and AP_M) and does not lift AP_S , confirming that confidence refinement cannot recover missed tiny plumes without better spatial evidence. In contrast, enabling the multiresolution representation produces the dominant improvement: AP_S increases from 0.037 to 0.188 and higher-IoU scores (AP_{75}) also rise, indicating sharper localization as well as recall gains. Combining multiresolution detection with the two-stage classifier yields the best overall performance ($AP_{50:95} = 0.410$), with the classifier acting as a lightweight stabilizer in the 0.1–0.3 band rather than as a primary source of recall. These results underline that the high-resolution composite is the key driver for faint-smoke sensitivity, while the selective second stage mainly curbs borderline false alarms without suppressing true positives.

In addition to accuracy, Table 5 also reports end-to-end inference time for each configuration. While the full pipeline introduces higher worst-case latency due to the additional multiresolution processing and the conditional second-stage classifier, this cost is not incurred uniformly across frames in practical surveillance settings. The classifier is applied only to detections within a narrow confidence band (e.g., 0.1–0.3), which occur infrequently given the rarity of early smoke events and the predominance of background-only frames. For example, over a stream of 10,000 frames, if the

classifier is triggered on only 10% of frames, the amortized per-frame runtime is approximately 9.5 ms, compared to 16.3 ms in the worst case. As a result, the most expensive component of the pipeline is triggered sparsely and in an event-driven manner, yielding an expected (amortized) runtime that remains close to that of the single-stage detector over long video streams. This design allows the system to prioritize sensitivity and robustness when ambiguous signals arise, while preserving efficient operation during normal background monitoring.

Table 5. Ablation study on the Early Smoke dataset evaluating the effects of multiresolution representation and the two-stage classifier. The classifier refines detections in the 0.1–0.3 confidence interval.

Multi-res	Classifier	AP _{50:95}	AP ₅₀	AP ₇₅	AP Small	AP Medium	AP Large	Time (ms)
X	X	0.345	0.760	0.264	0.037	0.236	0.436	3.168
X	O	0.343	0.768	0.260	0.037	0.246	0.424	7.515
O	X	0.404	0.813	0.378	0.188	0.313	0.475	5.825
O	O	0.410	0.820	0.383	0.188	0.318	0.482	11.708

5.4. Generalizability Verification

To assess generalizability, the proposed pipeline is evaluated on Pyro–SDIS under an identical operating threshold and refinement band. Table 6 shows that the proposed model achieves the highest AP@0.5:0.95, AP@0.5, and AP@0.75 among mainstream detectors, indicating that the multiresolution composite and skyline-guided placement transfer beyond the primary dataset. Scale-specific gains on small and medium objects suggest that the same mechanisms that enhance faint-plume sensitivity on Early Smoke also improve performance on Pyro-SDIS, while AP for size-specific objects remains competitive. This pattern indicates robustness to distribution shifts in background, scene layout, and smoke appearance without dataset-specific tuning.

Table 6. Comparative performance on the Pyro-SDIS dataset.

Model	AP@0.5:0.95	AP@0.5	AP@0.75	AP Small	AP Medium	AP Large
YOLOv5su	0.419	0.714	0.445	0.400	0.467	0.333
YOLOv8s	0.418	0.737	0.423	0.405	0.458	0.380
YOLOv9s	0.406	0.703	0.405	0.361	0.462	0.373
YOLOv10s	0.413	0.713	0.432	0.391	0.462	0.349
YOLOv11s	0.410	0.718	0.417	0.367	0.464	0.418
Ours	0.437	0.748	0.454	0.403	0.478	0.411

Table 7 reports deployment-setting performance on Pyro–SDIS using the same operating threshold ($Th_{conf} = 0.2$) and the same 0.1–0.3 refinement band. The multiresolution detector improves $AP_{50:95}$, AP_{50} , and AP_S over the baseline, and the selective two-stage pass adds incremental gains, reaching $AP_{50:95} = 0.415$, $AP_{50} = 0.700$, and $AP_S = 0.390$. The progression mirrors the primary dataset: multiresolution delivers most of the lift, and the classifier offers a modest stability boost, indicating that the deployment recipe transfers without dataset-specific tuning.

Table 7. Deployment-setting performance on the Pyro-SDIS dataset under a fixed decision threshold of $Th_{conf} = 0.2$. The classifier is applied only to detections within the 0.1–0.3 confidence band.

Multi-resolution	Classifier	AP _{50:95}	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L	Time (ms)
X	X	0.386	0.664	0.404	0.365	0.423	0.340	3.238
X	O	0.390	0.674	0.405	0.372	0.420	0.375	6.741
O	X	0.410	0.689	0.434	0.377	0.450	0.376	5.630
O	O	0.415	0.700	0.433	0.390	0.455	0.344	13.818

6. Conclusions

The proposed multi-resolution pipeline addresses the central challenge of detecting faint, small-scale wildfire smoke. By embedding a high-resolution crop within a standard square input and aligning it with skyline-guided placement, the detector recovers fine-grained plume structure while retaining global context. Ablation results show that this composite representation is the primary driver of the substantial AP gains on early-smoke instances, especially in AP_S, and that these benefits carry over to a secondary dataset without dataset-specific tuning. The selective two-stage classifier further stabilizes deployment behavior by refining detections only within a narrow uncertainty band, reducing false alarms near the operating threshold without altering scientific evaluation.

The pipeline delivers measurable improvements in recall and localization for early smoke while keeping latency within practical bounds for field deployment. The combination of single-pass multi-resolution detection, reversible coordinate remapping, and lightweight selective refinement provides a deployable balance between accuracy and efficiency for real-time wildfire monitoring. Future work will explore extending the same principles to temporal video cues and additional sensing modalities to further reduce false alarms under challenging environmental conditions.

Acknowledgments: This research was supported by the "Regional Innovation System & Education (RISE)" through the Seoul RISE Center, funded by the Ministry of Education (MOE) and the Seoul Metropolitan Government (2025-RISE-01-019-04). In addition, Tae-Hyuk Ahn was supported by the National Science Foundation (NSF) under Grant No. 2430236.

References

1. Arteaga, B.; Diaz, M.; Jojoa, M. Deep learning applied to forest fire detection. In Proceedings of the 2020 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE, 2020, pp. 1–6.
2. Robinne, F.N.; Secretariat, F. Impacts of disasters on forests, in particular forest fires. *UNFFS Background paper* **2021**.
3. Vasconcelos, R.N.; Franca Rocha, W.J.; Costa, D.P.; Duverger, S.G.; Santana, M.M.d.; Cambui, E.C.; Ferreira-Ferreira, J.; Oliveira, M.; Barbosa, L.d.S.; Cordeiro, C.L. Fire detection with deep learning: A comprehensive review. *Land* **2024**, *13*, 1696.
4. Zhao, Y.; Ma, J.; Li, X.; Zhang, J. Saliency detection and deep learning-based wildfire identification in UAV imagery. *Sensors* **2018**, *18*, 712.
5. Wu, H.; Li, H.; Shamsoshoara, A.; Razi, A.; Afghah, F. Transfer learning for wildfire identification in UAV imagery. In Proceedings of the 2020 54th Annual Conference on Information Sciences and Systems (CISS). IEEE, 2020, pp. 1–6.
6. Zhao, L.; Zhi, L.; Zhao, C.; Zheng, W. Fire-YOLO: a small target object detection method for fire inspection. *Sustainability* **2022**, *14*, 4930.
7. Ahn, Y.; Choi, H.; Kim, B.S. Development of early fire detection model for buildings using computer vision-based CCTV. *Journal of Building Engineering* **2023**, *65*, 105647.
8. Huang, P.; Chen, M.; Chen, K.; Zhang, H.; Yu, L.; Liu, C. A combined real-time intelligent fire detection and forecasting approach through cameras based on computer vision method. *Process Safety and Environmental Protection* **2022**, *164*, 629–638.
9. Girshick, R. Fast r-cnn. In Proceedings of the Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448.
10. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* **2018**.
11. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European conference on computer vision. Springer, 2020, pp. 213–229.
12. Lv, C.; Zhou, H.; Chen, Y.; Fan, D.; Di, F. A lightweight fire detection algorithm for small targets based on YOLOv5s. *Scientific Reports* **2024**, *14*, 14104.
13. Xiao, Z.; Wan, F.; Lei, G.; Xiong, Y.; Xu, L.; Ye, Z.; Liu, W.; Zhou, W.; Xu, C. Fl-yolov7: A lightweight small object detection algorithm in forest fire detection. *Forests* **2023**, *14*, 1812.
14. Zhang, L.; Wang, M.; Ding, Y.; Bu, X. MS-FRCNN: A multi-scale faster RCNN model for small target forest fire detection. *Forests* **2023**, *14*, 616.

15. Wang, Y.; Bashir, S.M.A.; Khan, M.; Ullah, Q.; Wang, R.; Song, Y.; Guo, Z.; Niu, Y. Remote sensing image super-resolution and object detection: Benchmark and state of the art. *Expert Systems with Applications* **2022**, *197*, 116793.
16. Keeley, J.E.; Syphard, A.D. Historical patterns of wildfire ignition sources in California ecosystems. *International journal of wildland fire* **2018**, *27*, 781–799.
17. Gill, A.M.; Stephens, S.L.; Cary, G.J. The worldwide “wildfire” problem. *Ecological applications* **2013**, *23*, 438–454.
18. Julian, K.D.; Kochenderfer, M.J. Distributed wildfire surveillance with autonomous aircraft using deep reinforcement learning. *Journal of Guidance, Control, and Dynamics* **2019**, *42*, 1768–1778.
19. Afghah, F. Autonomous Unmanned Aerial Vehicle Systems in Wildfire Detection and Management-Challenges and Opportunities. In Proceedings of the Dynamic Data Driven Application Systems, 2022.
20. Raczok, T.; Ivens, S.N.; Seidel, L.; Gehringer, S. Wildfire Detection and Monitoring: A Drone-Based Approach and Comparative Analysis. *Proceedings of the International ISCRAM Conference* **2025**.
21. Bailon-Ruiz, R.; Bit-Monnot, A.; Lacroix, S. Real-time wildfire monitoring with a fleet of UAVs. *Robotics and Autonomous Systems* **2022**, *152*, 104071.
22. Phillips Iii, W.; Shah, M.; da Vitoria Lobo, N. Flame recognition in video. *Pattern recognition letters* **2002**, *23*, 319–327.
23. Gubbi, J.; Marusic, S.; Palaniswami, M. Smoke detection in video using wavelets and support vector machines. *Fire Safety Journal* **2009**, *44*, 1110–1115.
24. Mambile, C.; Kaijage, S.; Leo, J. Application of deep learning in forest fire prediction: a systematic review. *IEEE Access* **2024**.
25. Ghali, R.; Akhloufi, M.A. *Remote Sensing* **2023**, *15*, 1821.
26. Akagic, A.; Buza, E. Lw-fire: A lightweight wildfire image classification with a deep convolutional neural network. *Applied Sciences* **2022**, *12*, 2646.
27. Seydi, S.T.; Saeidi, V.; Kalantar, B.; Ueda, N.; Halin, A.A. Fire-Net: A Deep Learning Framework for Active Forest Fire Detection. *Journal of Sensors* **2022**, *2022*, 8044390.
28. Ramos, L.; Casas, E.; Bendek, E.; Romero, C.; Rivas-Echeverría, F. Computer vision for wildfire detection: a critical brief review. *Multimedia Tools and Applications* **2024**, *83*, 83427–83470.
29. Li, J.; Tang, H.; Li, X.; Dou, H.; Li, R. LEF-YOLO: A lightweight method for intelligent detection of four extreme wildfires based on the YOLO framework. *International journal of wildland fire* **2023**, *33*, NULL–NULL.
30. Bhargav, R.; Singh, P. Efficient UAV-Based Forest Fire Detection Using CNN and YOLOv8 Integration. *2025 6th International Conference on Recent Advances in Information Technology (RAIT)* **2025**, pp. 1–6.
31. Jonnalagadda, A.V.; Hashim, H.A. SegNet: A segmented deep learning based Convolutional Neural Network approach for drones wildfire detection. *Remote Sensing Applications: Society and Environment* **2024**, *34*, 101181.
32. Khryashchev, V.V.; Larionov, R. Wildfire Segmentation on Satellite Images using Deep Learning. *2020 Moscow Workshop on Electronic and Networking Technologies (MWENT)* **2020**, pp. 1–5.
33. Bouguettaya, A.; Zarzour, H.; Taberkit, A.M.; Kechida, A. A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. *Signal Processing* **2022**, *190*, 108309.
34. Ghali, R.; Akhloufi, M.A.; Mseddi, W.S. Deep learning and transformer approaches for UAV-based wildfire detection and segmentation. *Sensors* **2022**, *22*, 1977.
35. Cheng, G.; Yuan, X.; Yao, X.; Yan, K.; Zeng, Q.; Xie, X.; Han, J. Towards large-scale small object detection: Survey and benchmarks. *IEEE transactions on pattern analysis and machine intelligence* **2023**, *45*, 13467–13488.
36. Wang, J.; Yang, W.; Guo, H.; Zhang, R.; Xia, G.S. Tiny object detection in aerial images. In Proceedings of the 2020 25th international conference on pattern recognition (ICPR). IEEE, 2021, pp. 3791–3798.
37. Jiang, L.; Yuan, B.; Du, J.; Chen, B.; Xie, H.; Tian, J.; Yuan, Z. MFFSODNet: Multiscale feature fusion small object detection network for UAV aerial images. *IEEE Transactions on Instrumentation and Measurement* **2024**, *73*, 1–14.
38. Hu, M.; Li, Z.; Yu, J.; Wan, X.; Tan, H.; Lin, Z. Efficient-lightweight YOLO: improving small object detection in YOLO for aerial images. *Sensors* **2023**, *23*, 6423.
39. Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors* **2023**, *23*, 7190.
40. Ozge Unel, F.; Ozkalayci, B.O.; Cigla, C. The power of tiling for small object detection. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2019, pp. 0–0.

41. Akyon, F.C.; Altinuc, S.O.; Temizel, A. Slicing aided hyper inference and fine-tuning for small object detection. In Proceedings of the 2022 IEEE international conference on image processing (ICIP). IEEE, 2022, pp. 966–970.
42. Zhang, H.; Hao, C.; Song, W.; Jiang, B.; Li, B. Adaptive slicing-aided hyper inference for small object detection in high-resolution remote sensing images. *Remote Sensing* **2023**, *15*, 1249.
43. Hao, C.; Zhang, H.; Song, W.; Liu, F.; Wu, E. Slinet: Slicing-aided learning for small object detection. *IEEE Signal Processing Letters* **2024**, *31*, 790–794.
44. Telçeken, M.; Akgun, D.; Kacar, S. An evaluation of image slicing and YOLO architectures for object detection in UAV images. *Applied Sciences* **2024**, *14*, 11293.
45. Muzammul, M.; Li, X.; Li, X. Enhancing Tiny Object Detection Using Guided Object Inference Slicing (GOIS): An efficient dynamic adaptive framework for fine-tuned and non-fine-tuned deep learning models. *Neurocomputing* **2025**, p. 130327.
46. Chen, Z.; Chen, G. STTSBI: A Fast Inference Framework for Small Object Detection in Ultra-High-Resolution Images. In Proceedings of the 2024 4th International Conference on Intelligent Technology and Embedded Systems (ICITES). IEEE, 2024, pp. 129–135.
47. Koyun, O.C.; Keser, R.K.; Akkaya, I.B.; Töreyin, B.U. Focus-and-Detect: A small object detection framework for aerial images. *Signal Processing: Image Communication* **2022**, *104*, 116675.
48. Wang, M.; Yue, P.; Jiang, L.; Yu, D.; Tuo, T.; Li, J. An open flame and smoke detection dataset for deep learning in remote sensing based fire detection. *Geo-spatial Information Science* **2025**, *28*, 511–526.
49. De Venâncio, P.V.A.; Rezende, T.M.; Lisboa, A.C.; Barbosa, A.V. Fire detection based on a two-dimensional convolutional neural network and temporal analysis. In Proceedings of the 2021 IEEE Latin American Conference on Computational Intelligence (LA-CCI). IEEE, 2021, pp. 1–6.
50. Team, P. Pyro-SDIS Dataset, 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.