Agentic Design for Wildfire Smoke Detection

Siena Okuno
University of California - San Diego
saokuno@ucsd.edu

Yuze Fu
University of California - San Diego
yuf042@ucsd.edu

Brighten Hayama
University of California - San Diego
bhayama@ucsd.edu

Tamoghno Kandar
University of California - San Diego
tkandar@ucsd.edu

1. Problem Statement

Climate change is driving warmer, drier conditions that have lengthened wildfire seasons and intensified the frequency and severity of wildfires across the United States. Given the serious threat these fires pose to people, property, and ecosystems, fast and early detection is critical to preventing catastrophic damage. Manual monitoring, however, is tedious and labor-intensive, motivating the need for automated solutions. To address this, we propose an agentic system for wildfire detection, moving beyond the limitations of traditional deep learning models.

2. Motivation

Modern wildfire detection systems typically involve the use of a CNN with a single image input to identify smoke. While CNNs excel at identifying visual patterns, false negatives and positives often arise due visual similarity between smoke and artifacts such as fog or clouds. Without important information like weather conditions or satellite fire alerts, these systems lack important context. This is exacerbated by the absence of temporal information.

An agentic framework with LLM integration can adaptively combine information from many different sources, such as visual outputs, weather data, and sensor readings. This allows the system to make more informed and reliable decisions. Additionally, because they can selectively use relevant tools instead of relying on a large, fixed model pipeline, agentic designs can lead to faster and more adaptive detection in real-world scenarios.

3. Related Work

With the rise in popularity of deep learning (DL), recent approaches to wildfire detection have moved away from rule-based methods to DL systems such as CNNs or LSTM[4–6, 10, 11]. These systems improve generalizability at the cost of requiring large quantities of labeled and representative data. Evaluation across methodologies remains inconsistent due to reliance on small, curated, or synthetic datasets, limiting real-world generalization and comparability[2].

Dewangan et al.[2] addressed the lack of data standardization with FIgLib, a labeled wildfire smoke detection dataset comprising sequences of wildfire images captured by fixed-view cameras on remote mountaintops in Southern California. Each sequence includes images from approximately 40 minutes before and after the start of a fire.

Govil et al. [3] evaluated an Inception V3 CNN [8] trained on a custom smoke dataset on FIgLib, but the authors reported false positives/negatives due to the confounding effect of fogs/clouds.

Dewangan et al. [2] proposed SmokeyNet, a computer vision pipeline that combined three types of neural networks,a CNN, an LSTM, and a vision transformer,to leverage spatiotemporal information from FIgLib image sequences for smoke detection. Although mitigation against false positives was taken, such as cropping the part of images that contain sky far above the horizon, the model still confused low hanging clouds with smoke[2].

Bhamra et al. [1] extended SmokeyNet with weather and satellite data. There was improvement to accuracy and time-to-detect when weather was included as a tensor input, but satellite imaging was found to be a weak predictor. This was potentially due to misalignment between spatial and temporal resolution between satellite images and FIgNet. The problem regarding false positive rate persisted.

Xie et al. [9] created the LLM agent WildfireGPT, which enriches an out-of-box LLM with domain-specific context to provide users with insight on wildfire-related conversations. However, the interface was text-only and not natively designed for detection.

Sandeep et al. [7] proposed an orchestrator-based multi-agent system for multimodal wildfire detection and insights. This model utilized a single unmanned aerial vehicle (UAV) image and text pair as input.

4. Expected Outcome

We aim to creaet an agentic wildfire detection system that addresses common false positives (misclassification of fog, low-hanging clouds, and haze) documented across SmokeyNet and similar models. The agentic system will dynamically incorporate multiple information sources and image sequences to make informed detection decisions. The goal is a reduction in false positive rate on ambiguous scenes with minimal sacrifices in recall or real-time latency, which we will evaluate against the existing baselines on the FIgLib dataset.

5. System Architecture

At a high level, our system can be decomposed into short-term memory, long-term storage, tools for agents to call, and the agents themselves. The agents are described in section 7.

5.1. Short-term Memory

5.1.1. Frame Buffer

The frame buffer points to the most recent image frames from a streaming camera, each of which has an update frequency of one frame per minute. The frame buffer allows the system to incorporate recent temporal context for smoke detection. For this implementation, we utilized ten frames per location and time pair in the evaluation set.

5.2. Long-term Memory

Long-term memory stores information needed during the initial wildfire risk and priority assessment. Data scalability is an important consideration as the ALERTCalifornia camera network includes thousands of locations. Because this information is only updated periodically, updates to long-term memory would ideally be run asynchronously and the results cached for later use. For the scope of this project, location and test date vary per evaluation sequence and thus location-specific information tools are called once at the beginning of every test sequence.

5.2.1. Local NCEI Weather Stations

Nearby weather stations are located for each camera location given an initial set of coordinates and a bounding box of 200 miles. Stations are sorted by distance from the target location and stored as a JSON containing station ID, station coordinates, available data types, and available data type date ranges.

5.2.2. Climate Summary

We store a summary of historical climate data pulled from local NCEI weather stations near the camera location. Up to five of the nearest stations are queried for their available datatypes. The Risk Assessment Agent determines relevant monthly datatypes, queries data from the last year, and stores it as an LLM-generated summary. As an example, some datatypes commonly referenced within the summary are average temperature, extreme maximum temperature, and precipitation (mm).

5.2.3. Weather Summary

Similar to climate data, historical weather data is also sourced from local NCEI weather stations. Due to station data upload schedules, the weather summary is limited to daily summary datatypes and not hourly updates. Some datatypes commonly referenced within the summary are average daily wind speed, average relative humidity for the day, and precipitation (mm).

5.3. Tools

5.3.1. NCEI API - Weather / Climate Information

This tool provides information about local NCEI weather stations. It can verify the availability of data types at each station or query stations for data. We estimate this tool to have minimum impact on runtime, but delegating it as an asynchronous task would remove latency altogether.

5.3.2. Farneback optical flow

This traditional optical flow tool provides initial smoke proposal points through estimate dense pixel wise motion between consecutive frames. However, traditional optical flow operates on the full image and may be sensitive to clouds, noise, unrelated background motion, or camera artifacts. We estimate this tool to have lower computational overhead than the SAM2 based optical flow.

5.3.3. SAM2 Point Tracking

SAM2 Point Tracking Tool allows the agent to segment and track potential smoke region acquired from Farneback optical flow tool across frames. Instead of estimating motion everywhere, this approach focuses on object level consistency by propagating segmentation masks over time. This allows the system to measure higher level temporal properties such as persistence, expansion, boundary diffusion, and structural changes of the plume. It is generally more robust to background noise but results depends on the quality of the initial segmentation. We estimate this tool to be more computational expensive.

5.3.4. SAM3 Prompt Tracking

SAM3 Prompt Tracking allows the agent to directly search for and track potential smoke regions using the text prompt "smoke." It gives the agent the option to detect smoke in cloudy or noisy frame sequences where traditional optical flow may be overwhelmed by unrelated motion. SAM3 generally provides better segmentation and tracking quality, but it is the most computationally expensive tool.

5.3.5. Siglip2

This tool provides additional global image context regarding cloudy versus clear days on top of Gemma's native scene understanding and passed risk context.

6. Representation

Agent communication is enabled by textual representation throughout the detection pipeline. Depending on the process step, this will contain information of the following categories: scene context, risk context, detection state, and loop control state. Not all fields are visible to all agents.

Scene context includes the camera ID, timestamp, location, and a reference to the frame buffer. Timestamp and location are extracted from the camera feed via OCR.

Risk context is updated for each camera's local area. It is comprised of the wildfire risk score, area priority, and uncertainty factors. The wildfire risk score describes the likelihood of a fire to start within an area and is rated as low, medium, or high. The area priority is rated similarly and is an indicator for the likelihood of a fire to become out of control or be abnormally destructive. Uncertainty factors describe informed predictions surrounding the risk of false positives. For example, predicted haze due to a locality's dry climate increases the risk of false positives.

The detection state captures the wildfire smoke detection label, confidence score, bounding region (if smoke is present), and chain-of-thought trace for the tool calls made.

The loop control state holds information on output critiques, recent confidence scores, and current iteration count. It is used for planning to determine whether to terminate the decision-critique loop, re-task agents with updated context, or escalate to human review.

7. Agent Role

The system is divided into four specialized roles: planning, risk assessment, primary smoke detection, and verification. This separation makes the pipeline more reliable because each agent focuses on a distinct part of the reasoning process. The Primary Detection Agent focuses on visual evidence, the Risk Assessment Agent provides environmental context, the Critic Agent evaluates reliability, and the Manager/Planner Agent coordinates the overall workflow.

This agentic design is intended to reduce false positives and false negatives. False positives may occur when clouds, fog, dust, glare, or camera artifacts resemble smoke. False negatives may occur when early-stage smoke is subtle, partially occluded, or only visible across multiple frames. By combining visual detection, temporal reasoning, environmental risk assessment, and independent critique, the system attempts to make more robust decisions than a standalone classifier.

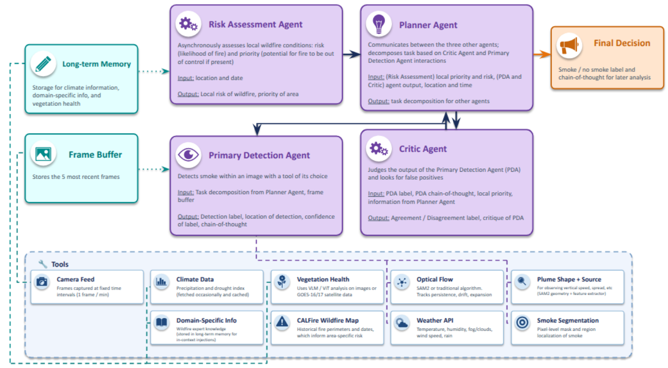

Figure 1. Diagram of agentic workflow

7.1. Planner Agent

The Planner Agent acts as the central coordinator of the workflow. It decomposes the smoke detection task, routes information to the appropriate agents, assesses task complexity and ambiguity, and determines whether the workflow should continue or terminate. Its inputs include outputs from the Risk Assessment Agent, Primary Detection Agent, and Critic Agent, along with the current camera frame, system time, and any OCR-extracted timestamp or location information.

A key responsibility of the planner is deciding when the system has reached a reliable conclusion. If the Primary Detection Agent and Critic Agent agree with sufficient confidence, the planner can terminate the reasoning loop for the current timestep. If they disagree, it may request processing of additional frames, trigger stronger models, call for extra verification tools, continue the loop, or decide human verification is required.

The planner also controls what information each agent receives. For the Primary Detection Agent, it provides only task-relevant information such as the desired confidence level, time, location, and scene background. It does not provide weather conditions, wildfire risk, or area priority, so that the primary prediction remains grounded in visual evidence. In contrast, the Critic Agent is given richer context such as weather, area priority, and estimated wildfire risk, allowing it to evaluate the plausibility of the primary result more broadly.

As future work, the planner would ideally maintain memory of recent indecision incidents and track larger uncertainty trends to adapt the workflow.

7.2. Risk Assessment Agent

The Risk Assessment Agent estimates the local wildfire risk and priority level for the monitored camera region. Since environmental conditions change more slowly than visual smoke patterns, this agent does not need to update with every frame. Instead, it would ideally be run periodically using inputs such as camera location, current date, and available metadata.

Its output is the risk context: local wildfire risk, area priority, and the likelihood of false positives under current conditions. For example, a dry and vegetation-heavy region near an urban area during fire season would receive a higher risk and priority score than a barren and remote location. This information is consumed by the Planner and Critique Agents, but is not passed to the Primary Detection Agent so as to avoid bias.

The Risk Assessment Agent uses the NCEI API tool to summarize local climate and weather, which is ingested at detection runtime to determine the risk context. As future work, the agent could be extended with historical wildfire perimeter information, locality-specific wildfire retrieval tools, and could help adjust the camera sampling rate based on wildfire risk and area priority.

7.3. Primary Detection Agent

The Primary Detection Agent detects smoke in camera imagery. Its main input is the temporal frame buffer. This allows the agent to reason about both the appearance and behavior of a suspected smoke region. Temporal evidence is important because smoke often drifts, expands, or changes shape, while clouds, haze, glare, or static artifacts may behave differently.

To allow scalability, the Primary Detection Agent uses the expensive SAM3 text prompt for smoke proposal/segmentations/centroids tracking if own scene understanding over the frame buffer or risks context from previous agents contain clouds/fog/hazes. Otherwise, the Primary Detection Agent uses the cheaper optical flow for initial smoke point proposals and acquire segmentation and centroids tracking through SAM2 point prompt.

The Primary Detection Agent distinguish clouds or other atmospherical artifacts against smoke through reasoning over smoke segmentations area changes and centroid movements.

Its output is the detection state part of the system's representation: a smoke detection label, confidence score, and the location of the suspected smoke region if smoke is detected. It also provide a concise methodology summary describing the visual evidence and tools used, rather than relying on hidden reasoning.

7.4. Critic Agent

The Critic Agent evaluates the reliability of the Primary Detection Agent's output. It receives the full detection context of detection label, confidence score, predicted smoke location, tool-call chain, and methodology summary. It also receives risk context information to increase or lower the harshness of its judgement.

Unlike the Primary Detection Agent, the Critic Agent is allowed to use broader environmental context. Its role is not to make the initial visual prediction, but to judge whether the prediction is reliable and whether the primary agent used an appropriate method. For example, if the primary agent reports smoke with high confidence but skips temporal analysis or plume verification in an ambiguous scene, the critic may flag the result as unreliable.

The Critic Agent outputs whether it agrees or disagrees with the primary label. It also critiques the methodology and can recommend rerunning detection with stronger models or additional verification tools. The Manager/Planner Agent then uses this feedback to decide whether to terminate, request more evidence, increase model complexity, or escalate to human review.

8. Method Overview

8.1. Compute Resources

LLM inference is provided by NRP Nautilus. Local CV tool inference relies on a single NVIDIA RTX 4080.

8.2. Test Methodology

We evaluate our agentic system on a total of 40 FIgLib sequences sampled between 2020-2025, each with 10 frames near the start of the fire. Upon manual inspection, we find on average 8 of the sequences to not contain any detectable smoke. We further categorize the test sequences into *complex* or *simple* cases. Here, we define a video sequence to be complex if it contains one or more of SmokeyNet's typical failure modes: low-hanging or fast-moving clouds, haze, glare, camera artifacts, etc. A test sequence is simple if the sky is clear and view of the smoke is unobstructed. Out of the 40 test sequences, we find 24 to be complex and 16 to be simple.

8.3. Prediction Label Evaluation

The classification output of our workflow is evaluated across three metrics: accuracy, recall, and time-to-detect.

8.3.1. Accuracy and Recall

Accuracy and recall are vital to evaluate the automated detection of wildfires. False positives can cause fatigue for human staff who verify smoke prediction results and allocate resources, but small rates of false positives are tolerated. False negatives can be catastrophic in regards to both immediate harm as well as harm caused by a fire becoming out of control.

8.3.2. Time-to-detect

First-responders and wildfire monitoring personal at CAL-Fire aim for a sub-eight minute response time for wildfires, including mobilization time. Incoming camera images are sampled at a rate of one image per minute and thus both inference time and the number of frames required to detect are vital. As our test data is retrieved from archived examples instead of live camera feeds, we are limited in our ability to accurately report time-to-detect (TTD). For simplicity, we only report the runtime in seconds from the beginning of the risk assessment task to the final decision.

8.4. Priority Assessment Evaluation

Both risk score and area priority are reference-free evaluation metrics, but important to ground in reality. As future work, results should be compared to CALFire's Fire Hazard Severity Zone (FHSZ) database, which defines a hazard score based on the factors that influence fire likelihood and behavior. FHSZ hazard scores are limited to estimates of wildfire probability, not impact. An area can have a hazard score of none, moderate, high, or very high.

9. Implementation Details

The agentic system is an offline, auditable wildfire smoke detection workflow over image frame buffers built with a CrewAI framework. The implemented pipeline runs once per ten input camera frames. It is implemented in a sequential manner for simplicity, but is able to run up to five loops.

The Risk Assessment Agent is split into two distinct steps; it summarizes the climate and weather and then assesses risk and priority. This seperation is to prevent the explosion of the agent context with its many tool calls, some of which load the same data with different goals. After creating summaries, it generates the full risk context of `risk`, `priority`, and `uncertainty` to pass to the Planner Agent.

The planner agent sets `CONFIDENCE_THRESHOLD`: high risk starts at 0.85, medium risk at 0.70, and low risk at 0.55. This threshold controls how closely the Primary Detection Agent and Critic Agent should agree before the final decision is considered reliable. The planner may raise the threshold when the loop produces cloud or fog false positives, or lower it when smoke is likely to be missed. It also selects the initial vision route (SAM3 or traditional optical flow) based on haze, clouds, fog, marine layer, or broad horizon.

The Critic Agent receives the local weather and climate summaries, risk assessment, planner instructions, and Primary Detection Agent output, then checks whether the reported label is supported by visual evidence, source localization, and temporal metrics. When SAM3 finds a smoke candidate, the critic checks whether the optical-flow centroid analysis was performed and whether the resulting motion is compact. Critic disagreement is passed forward as an audit signal for the final decision, though we observe in our results that this often leads to many false-negatives.

The vision path uses SAM3 text-prompt segmentation, or traditional optical flow when appropriate, to produce initial smoke candidates; SAM2 then tracks temporal persistence, area growth, and centroid drift across the frame sequence. Rather than treating a segmentation match as a detection, the agent reasons over the resulting centroid-movement and area-growth signals to separate smoke from clouds or marine-layer fog. The working assumption is that SAM3 text-prompt segmentation outperforms the traditional optical-flow and SAM2 path, and that the centroid-movement / area-growth signatures of cloud and smoke objects are distinct.

The final decision policy requires both visual plausibility and temporal plume behavior; growth or drift alone is not treated as sufficient evidence when a cloud or marine-layer explanation is plausible.

9.1. Model Parameters

Because the agentic workflow is dynamic, the total number of model parameters used in a given run depends on which tools are invoked and how many reasoning steps are required. Table 1 lists the parameter counts of the main models available to the system. This separates the static model sizes from the run time compute path, allowing us to report a minimum workflow cost while still accommodating more complex cases that require additional tool calls.

Table 1. Parameter counts for the models used in our system.

Model	Parameter Count
Gemma4	31B
SAM3	0.9B
Siglip2	0.4B
SAM2 Hiera Large	0.2B

9.1.1. Minimum workflow

In the minimum workflow, the system performs six Gemma4 LLM invocations and one SAM2 Hiera Large invocation. Under this lower bound path, the cumulative parameter usage is

$$6 \times 31\text{B} + 1 \times 0.2\text{B} = 186.2\text{B}.$$

This value should be interpreted as the minimum cumulative parameter count across model calls, rather than the number of unique parameters stored in memory. Additional calls to Gemma4, SAM2, SAM3, or SigLIP2 increase this total according to the tools selected by the agents during execution.

10. Results

10.1. Final Decision Label Results

Table 2 reports performance by scene difficulty. We observe perfect precision across the 40 test sequences, but only detect smoke in half of the overall smoke sequences. Our agentic system struggles more with complex scenes, possibly due to the small and transparent appearance of early-stage smoke plumes within the 10-frame window, high variability in plume growth, and frequent confounders like low-hanging clouds or haze.

10.2. Risk Assessment Agent Results

10.2.1. Climate and Weather Summaries

The climate and weather summaries were well-grounded in NCEI weather station data and explicitly referenced these in their summaries. Clearing the context before each summary allowed the agent to correctly assess the available data and make informed decisions regarding which to include in each summary. Commonly referenced datatypes for climate were average temperature, extreme maxium temperature, and precipitation. Weather summaries often mentioned average daily wind speed, average relative humidity, precipitation, and average temperature. Consistent references to extreme temperatures and dry precipitation patterns are consistent with the California Palmer Drought Severity Index at the time of our evaluation data.

The summaries consistently mentioned an increased likelihood of false positives during smoke detection due to potential haze or clouds, where haze is due to dry conditions. These likely informed the uncertainty output of the risk assessment and may have biased the overall workflow to be overly-cautious of false postives. Given the importance of recall over precision, this would need to be addressed in future iterations.

A notable limitation was the number of NCEI weather stations within California. While NCEI does not explicitly list the number of available stations, many summaries traced their reasoning to less than five stations. There were also mentions of uncertainty due to the difference in elevation or distance from the weather stations and the target location.

Analysis was not performed on the tool-chain call in regards to the usage frequency of datatypes. As future work, it would likely be beneficial to identify key datatypes that were largely overlooked and perform additional prompt tuning to include these. Also, tracking usage and pre-querying common data could prevent the need to call an LLM for relevant datatype analysis. This could reduce cost or system computation requirements.

10.2.2. Wildfire Risk and Local Priority

The wildfire risk and local priority were typically both reported as ”High”. The risk score is consistent with the expected assessment, as our evaluation data is all from times of a wildfire incident or immediately preceeding one. The largest contributing factors to risk assessment were aggregated weather and climate information related to drought and temperature.

Local priority is primarity determined by average wind speeds from the previous week, though uncertainty is noted in regards to current wind speed due to a lack of data.

10.2.3. Uncertainty

Uncertainty is consistently mentioned in regards to a lack of real-time data for wind speed, fog/cloud conditions, and historical fire perimeters. Wind speed and historical fire perimeters inform priority; winds cause fires to spread rapidly while historical fire perimeters provide information on the quantity of fuel within an area. Fog/cloud conditions inform the likelihood of false positives.

10.3. Planner Agent Results

The Planner Agent generally translated the Risk Assessment Agent output into conservative downstream behavior. The planner typically mentioned SAM3 and included cloud/fog false-positive checks. In these cases, the planner instructs the Primary Detection Agent to begin with SAM3 when atmospheric confounders are plausible, and to follow any returned smoke candidate with optical-flow centroid analysis. Because uncertainty from the risk assessment often mentioned fog, haze, or limited real-time weather data, the planner frequently emphasized false-positive prevention and close agreement between the Primary Detection Agent and Critic Agent. This likely contributed to the perfect precision but low recall. The planner output was also relatively long, which may have caused issues due to the context-rot phenomenon. Future iterations should make risk and priority more explicitly metric-based and shorten planner instructions so that cloud/fog caution does not dominate early-smoke detection.

10.4. Detection Agent Results

10.4.1. Centroid Analysis

When potential smoke is segmented and tracked by either the SAM2 or SAM3 models, the Primary Detection Agent demonstrates the ability to distinguish smoke, whose centroid drift and cumulative motion remain relatively upward and compact, from clouds that exhibit broad layer wise movement. This reasoning works well when smoke does not form a horizontal inversion layer near ground level.

10.4.2. Gemma's Vision Understanding

The Primary Detection Agent's underlying model is Gemma4. Smoke segmentations that pass centroid analysis are checked using Gemma's vision capability to assess source location. Based on the agent output logs during evaluation, Gemma's scene understanding may have demonstrated an ability to determine whether suspected smoke segments originate from specific points in the terrain or vegetation. However, more testing is needed to draw a conclusive claim about Gemma's ability to understand the source of potential smoke segmentations.

Table 2. Prediction results on the 40-sequence evaluation set

Split	Sequences	Precision	Recall	Accuracy
All	40	1.00	0.50	60.0%
Complex	24	1.00	0.43	50.0%
Simple	16	1.00	0.64	75.0%

10.5. Critic Agent Results

The Critic Agent generally acted as an effective false-positive filter. It frequently challenged smoke candidates that looked like broad atmospheric layers, especially when SAM3 returned large horizon-spanning masks or when optical-flow centroid analysis showed broad layer-wise drift rather than compact plume motion. In successful cases, this allowed the workflow to reverse an initial SAM3 ”smoke” segmentation into a final `no_smoke` decision for low-hanging clouds or marine-layer-like objects.

However, the critic also appears to contribute to the system's conservative bias. Because it was instructed to aggressively check for fog, haze, clouds, missing temporal evidence, and weak source localization, it sometimes questioned early smoke detections when the plume was small, distant, or appeared near the end of the 10-frame window. This likely helped maintain perfect precision, but may have reduced recall by making the final decision overly dependent on strong temporal evidence.

10.6. Case Study: Confounder Rejection

The clearest illustration of our system's perfect precision is the Otay Mountain West sequence (Border Fire, 2020-08-28), a non-smoke scene with low-hanging clouds that has induced false positives in prior work.

SAM3 text-prompt segmentation initially masked the horizontally moving cloud as “smoke.” Further optical-flow analysis revealed a broad-region signal with large centroid drift (399.53 px) and no compact, localized motion. VLM-based (Gemma) source localization found no terrain or vegetation point-source, and the agent reversed the SAM3 decision, labeling the sequence `no_smoke`.

This highlights the capabilities of our agentic design, incorporating more nuanced agent reasoning and combined tool-use in the overall detection label.

10.7. Failure Modes

10.7.1. Distant or transparent smoke

Faint, distant, or semi-transparent plumes cause SAM3 to return no candidate region, which propagates to a `no_smoke` label. This is the primary reason for the low recall in complex scenes. Potential future directions include recovering these candidates with the traditional optical-flow + SAM2 path on clear days, finding an alternative to SAM3 on cloudy days, and using optical flow to crop or anchor the search region.

Another option is to experiment with the number of frames included during detection inference time. Small smoke plumes could be overly diluted within our current context, and thus reducing the included frames may prove beneficial.

10.7.2. Late ignition

When a fire begins near the end of the ten-frame window, the optical-flow reasoning chain has too little temporal evidence and the agents discount an otherwise valid SAM3 detection.

A future direction to address this pattern is allowing Primary Detection Agents to access historical frames for the camera location. Such explicit temporal information may enable the agent to reason if the newly detected potential smoke is connected from any specific terrain or vegetation point sources, detect anomolies, etc.

10.7.3. Inversion Layers

Certain climates and weather conditions can make smoke travel horizontally and breaks the low hanging clouds versus smoke reasoning that the Primary Detection Agent currently relies on.

Through applying domain knowledge and adding satellite temperature data, we can make the reasoning regarding smoke segmentation movement and growth aware of when horizontal smoke is likely. This can prevent the critic or primary detection agents from misclassifying inversion layers.

10.7.4. Harsh critic

The risk assessment's caution about haze and clouds propagates into harsh critic behavior. This, along with potential SAM3 failures in complex cases, results in false negatives.

10.7.5. Runtime

The 200-300s loop is dominated by LLM calls. A mixture-of-experts or smaller agentic model could reduce this, though we have not explored it.

References

[1] Jaspreet Kaur Bhamra, Shreyas Anantha Ramaprasad, Siddhant Baldota, Shane Luna, Eugene Zen, Ravi Ramachandra, Harrison Kim, Chris Schmidt, Chris Arends, Jessica Block,

Ismael Perez, Daniel Crawl, Ilkay Altintas, Garrison W. Cottrell, and Mai H. Nguyen. Multimodal wildland fire smoke detection. *Remote Sensing*, 15(11), 2023. 1

[2] Anshuman Dewangan, Yash Pande, Hans-Werner Braun, Frank Vernon, Ismael Perez, Ilkay Altintas, Garrison W. Cottrell, and Mai H. Nguyen. Figlib & smokeynet: Dataset and deep learning model for real-time wildland fire smoke detection. *Remote Sensing*, 14(4), 2022. 1

[3] Kinshuk Govil, Morgan L. Welch, J. Timothy Ball, and Carlton R. Pennypacker. Preliminary results from a wildfire detection system using deep learning on remote camera images. *Remote Sensing*, 12(1), 2020. 1

[4] Mira Jeong, MinJi Park, Jaeyeal Nam, and Byoung Chul Ko. Light-weight student lstm for real-time wildfire smoke detection. *Sensors*, 20(19), 2020. 1

[5] Tingting Li, Enting Zhao, Junguo Zhang, and Chunhe Hu. Detection of wildfire smoke images based on a densely dilated convolutional network. *Electronics*, 8(10), 2019.

[6] Minsoo Park, Dai Quoc Tran, Daekyo Jung, and Seunghee Park. Wildfire-detection method using densenet and cyclegan data augmentation-based remote camera imagery. *Remote Sensing*, 12(22), 2020. 1

[7] Ashen Sandeep, Sithum Jayarathna, Sunera Sandaruwan, Venura Samarappuli, Dulani Meedeniya, and Charith Perera. Context-aware multi-agent architecture for wildfire insights. *Sensors*, 26(3), 2026. 1

[8] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016. 1

[9] Yangxinyu Xie, Bowen Jiang, Tanwi Mallick, Joshua David Bergerson, John K. Hutchison, Duane R. Verner, Jordan Branham, M. Ross Alexander, Robert B. Ross, Yan Feng, Leslie-Anne Levy, Weijie Su, and Camillo J. Taylor. Wildfiregpt: Tailored large language model for wildfire analysis, 2025. 1

[10] Qi xing Zhang, Gao hua Lin, Yong ming Zhang, Gao Xu, and Jin jun Wang. Wildland forest fire smoke detection based on faster r-cnn using synthetic smoke images. *Procedia Engineering*, 211:441–446, 2018. 2017 8th International Conference on Fire Science and Fire Protection Engineering (ICFSFPE 2017). 1

[11] Feiniu Yuan, Lin Zhang, Xue Xia, Boyang Wan, Qinghua Huang, and Xuelong Li. Deep smoke segmentation. *Neurocomputing*, 357:248–260, 2019. 1