

## AI 기반 불꽃과 연기 동시 탐지 기술을 통한 산불 조기 탐지 적용 방안

류재영 \* · 박문환 · 강우성 · 강수형  
데이터쿡

### AI-Based Flame and Smoke Dual-Detection Technology for Early Wildfire Detection

JaeYoung Ryu \*, Munhwan Park, WooSeong Kang and SooHyung Kang

DataCook, Seoul 04034, Korea

**요약:** 최근 지구 온난화의 가속화로 인해 산불의 발생 빈도와 강도가 전 세계적으로 증가하고 있으며, 이로 인한 온실가스 및 대기 오염물질의 배출이 기하급수적으로 증가하고 있다. 산불이 대형화될수록 탄소 배출, 대기질 악화, 생태계 파괴, 건강 피해 등 사회적·환경적 영향이 악화될 것으로 예측되고있으며, 이러한 문제점을 다소나마 해결하기 위하여 산불을 조기에 탐지할 수 있는 인공지능(AI) 기술, 특히 딥러닝과 영상 분석을 적용한 다양한 연구가 활발히 진행되고 있다. 그러나 기존의 영상 기반 화재 감지 기술은 불꽃(flame) 또는 연기(smoke) 중 하나의 단일 신호에만 의존하거나, 비교적 후기 단계에서 화재를 탐지하는 방식이 주를 이루어, 초기 단계에서의 시각적 징후를 포착하기 어렵다는 한계를 가진다.

본 연구에서는 이러한 문제를 해결하기 위하여 불꽃과 연기를 동시에 탐지할 수 있는 AI 기반 복합 감지 프레임워크를 제안한다. 본 연구에서는 연기와 불꽃 감지에 특화된 모델을 복합적으로 사용하여 미약한 초기 신호를 효과적으로 인식할 수 있도록 최적화된 하이브리드 딥러닝 구조를 적용하였다. 실험 결과, 본 연구에서 제안한 연기·불꽃 동시 감지 모델은 기존 단일 영상(신호) 기반 탐지 방식에 비해 산불 초기 단계에서의 탐지 정확도와 환경 일반화 성능이 유의미하게 향상됨을 확인하였다. 본 연구의 결과는 실시간 산불 감시 및 예방 시스템의 핵심 기술로 활용될 수 있는 가능성을 제시한다.

**Abstract:** The frequency and intensity of wildfires are increasing worldwide due to the accelerating pace of global warming, leading to an exponential rise in greenhouse gas emissions and air pollution. As wildfires become more severe, their social and environmental impacts, such as carbon emissions, air quality deterioration, ecosystem destruction, and health hazards, are expected to worsen dramatically. To mitigate these challenges, various studies have actively explored artificial intelligence (AI) technologies, particularly deep learning (DL) and computer vision, for early wildfire detection. However, conventional vision-based fire detection systems typically rely on a single signal, either flame or smoke, or focus on detecting fires in later stages, making it difficult to capture subtle visual cues during the early ignition phase.

To address these limitations, this study proposes an AI-based dual-detection framework that can simultaneously identify flame and smoke signals. The proposed framework employs a hybrid DL architecture that combines specialized flame and smoke detection models to effectively recognize faint early-stage signals. Experimental results demonstrate that the dual-detection model substantially improves detection accuracy and environmental generalizability compared with conventional single-signal detection models. The findings suggest that the proposed framework has strong potential to serve as a core technology for real-time wildfire monitoring and prevention systems.

**Key words:** wildfire, detection, flame, smoke, smokeynet, yolov8, ensemble, response

\* Corresponding author  
E-mail: rjy@datacook.kr

ORCID  
JaeYoung Ryu  https://orcid.org/0009-0002-2143-1566

## 서론

최근 지구 온난화가 가속됨에 따라 산불 발생 빈도도 증가하는 추세이며, 이는 온실가스와 오염물질의 배출을 촉진시키고 있다. 특히 대형 산불이 더 빈번하고 강력해질수록, 이로 인한 탄소 배출과 공기 오염, 건강 피해가 더욱 확대될 것으로 예측되고 있다(Zhao et al., 2025). 이에 산불의 조기 감지를 목표로 하는 기계학습 및 딥러닝 기반 인공지능 기술에 대한 관심이 높아지고 있다(Wasserman and Mueller, 2023). 기존 영상 기반 화재 감지 기술은 주로 불꽃(flame)이나 연기(smoke) 중 하나의 단일 신호에 집중하거나, 비교적 후기 단계의 화재를 탐지하는 데 초점을 맞춰 왔다. 이러한 접근은 초기 단계의 약한 시각적 징후를 포착하기 어렵고, 다양한 크기·환경 조건에서 일반화 성능이 낮다는 한계를 가진다(Part et al., 2016). 또한 환경 요인에 따른 비화재 오탐(false alarm) 문제 역시 해결되지 않은 주요 과제로 지적되고 있다(AI-Smadi et al., 2023) 구름, 안개 등은 실제 연기와 매우 유사한 시각적 패턴을 생성하여 모델의 오류를 유발한다. 기존 연구들에서도 이러한 환경적 요인으로 인해 연기(smoke) 클래스의 precision 저하가 반복적으로 보고되어 왔다. 더불어 단일 모델 기반 접근은 특정 데이터셋의 특성이나 촬영 조건에 과도하게 종속되는 경향이 있어 안정적인 화재 탐지 성능을 유지하기 어렵다. 이러한 문제를 해결하기 위해 본 논문에서는 연기와 불꽃을 동시에 감지할 수 있는 탐지 기법을 제안한다.

## 연구 방법

### 1. 데이터셋 구성

#### 1) 데이터 소스 및 수집

본 연구에서는 3가지 주요 데이터 소스를 활용하여 학습 데이터셋을 구축하여 학습을 진행하였다.

#### (1) Roboflow 오픈소스 데이터셋

Table 1과 같이 초기 학습 데이터는 Roboflow 플랫폼을 통해 공개된 6개의 화재 탐지 관련 데이터셋을 통합하여 구성하였다. 이들 데이터셋은 Fire and Smoke Dataset, fire and smoke detection Dataset, Fire detection v3 Dataset, wcnm Dataset, Smoke\_4 Dataset, FDS Dataset으로 구성되며, 다양한 촬영 각도, 조명 조건, 거리에서 수집된 실제 화재 이미지와 바운딩 박스 형태의 객체 탐지 어노테이션을 포함한다.

#### (2) AI Hub 화재 3D 데이터셋

Table 2와 같이 과학기술정보통신부와 한국지능정보사회진흥원(NIA)이 주관하는 AI Hub의 “화재 3D 영상” 데이터셋을 활용하였다. 본 데이터셋은 실내외 화재 상황을 3D 형태로 구축한 고품질 데이터로, 다양한 각도와 시점에서의 화재 영상을 포함하며, 특히 초기 화재 단계의 미약한 불꽃과 연기 신호를 포착하는 데 유용하다.

**Table 1 Roboflow dataset composition.**

Data	Subset	Number of Data	Ratio
Fire & Smoke Data	TEST	3,216	4.97%
	TRAIN	54,271	83.90%
	VALIDATION	7,195	11.12%
	TOTAL	64,682	100.00%

**Table 2. Fire 3D dataset composition.**

Data	Type	Description	File Format	Quantity
Fire Video 3D Object Data	Raw Data	FDS input files	fds	46
	Raw Data	Plot3D	plot3D	28,051
	Source Data	Simulation result videos	mp4	138
	Source Data	Fire 3D object data	fga	2,440
Indoor 3D Space Data	Raw Data	LiDAR scan data	LAS, E57	93
	Raw Data	Indoor space photos (2D images)	jpg	932
	Source Data	Indoor 3D space modeling data	fbx, max, mb	78
	Source Data	Object texture files	.fbx, .png, .jpg, .tga, .obj, .tx, .tif, .ies, .dds, .jiff, bmp, gif	27,894
Fire Video 2D Image Data	Source Data	Fire/Smoke images	jpg	153,930
Total				213,602

(3) AI Hub 지역안전재난(산불) 방재 데이터셋

Table 3과 같이 클래스 불균형 문제를 해결하기 위해 AI Hub의 “지역안전재난(산불) 방재의 고도화를 위한 대규모 인공지능 데이터베이스 구축” 데이터셋을 추가로 수집하였다. 본 데이터셋은 한국임업진흥원과 국립산림과학원이 구축한 것으로, 국내 산림 지역의 실제 산불 현장 및 시뮬레이션 데이터를 포함한다. 산불 초기 단계의 불꽃 이미지, 다양한 농도와 거리의 산불 연기, 산악 지형 특성이 반영된 화재 데이터, 시간대별(주간/야간), 계절별, 기상 조건별 데이터를 포함하며, 불꽃(Fire)과 연기(Smoke) 클래스별 바운딩 박스 정보가 전문가 검수를 거쳐 제공된다.

2) 초기 모델 학습 및 클래스 불균형 문제 발견  
초기 Roboflow과 AI Hub 화재 3D 데이터셋을 활용한

YOLO 모델 학습 결과, 불꽃 탐지 평균정확도(mAP50)는 0.977, 연기 탐지 평균정확도는 0.684, 전체 객체 탐지 평균정확도는 0.830으로 나타났다. 여기서 mAP50은 두 바운딩 박스가 겹치는 비율(IoU) 50% 이상인 예측값 대비 실제 검출(True positive)된 클래스의 평균정확도이다. 불꽃 탐지 정확도가 연기 탐지 정확도에 비해 상대적으로 높게 나타난 것은 학습데이터의 수량이 적어 모델 학습 시 과적합(overfitting)이 발생한 것으로 추정되었다. 데이터 분석 결과, Figure 1과 같이 연기(Smoke) 클래스의 샘플 수가 불꽃(Fire) 클래스에 비해 압도적으로 많은 클래스 불균형 현상이 확인되었다.

이러한 불균형은 모델이 다수 클래스에 편향되어 학습되는 문제를 야기하였다.

Table 3. Regional safety & disaster (Wildfire) prevention dataset composition positive data.

Data	Image Type	Collection Method	Object Classification	Collection Quantity	Subtotal	
Positive Data	Positive Real DB	Ground	Flame	57,600	360,000	
			Black Smoke	76,800		
			White/Gray smoke	153,600		
		Drone	Flame	14,400		
			Black Smoke	19,200		
			White/Gray smoke	38,400		
	Positive Synthetic DB	Ground	Flame	22,400	140,000	
			Black Smoke	29,850		
			White/Gray smoke	59,750		
		Drone	Flame	5,600		
			Black Smoke	7,450		
			White/Gray smoke	14,950		
Subtotal			Flame	100,000	500,000	
		Black Smoke	133,300			
		White/Gray smoke	266,700			
Negative Data	Negative DB	Ground	Cloud	64,000	200,000	
			Chimney smoke	64,000		
			Fog/Haze	32,000		
		Drone	Cloud	16,000		
			Chimney smoke	16,000		
			Fog/Haze	8,000		
	Subtotal			Cloud	80,000	200,000
			Chimney smoke	80,000		
			Fog/Haze	40,000		
Grand Total		Ground		560,000	700,000	
		Drone		140,000		
		Subtotal		700,000		

3) 클래스 불균형 문제 해결

클래스 불균형 문제를 해결하기 위해 먼저 소수 클래스인 불꽃 데이터를 보완하고자 추가 데이터 수집을 시도하였으나, 실제 불꽃 이미지의 희소성으로 인해 충분한 데이터 확보에 한계가 있었다. 이에 AI Hub의 “지역안전재난(산불) 방재의 고도화를 위한 대규모 인공지능 데이터베이스 구축” 데이터셋을 통해 불꽃 데이터를 추가 수집하였다. 추가 데이터 수집 이후에도 여전히 클래스 간 불균형이 존재하여, Figure 2와 같이 다운샘플링(Down-sampling) 기법을 적용하였다.

Figure 3과 같이 다수 클래스인 연기 데이터의 샘플 수를 소수 클래스인 불꽃 데이터 수준으로 감소시켜 클래스 간 균형을 맞추었다.

다운샘플링 과정에서는 다양한 환경 조건을 대표할 수 있는 샘플을 우선 선정하고, 중복되거나 유사한 이미지를 제거하며, 어노테이션 품질이 우수한 데이터를 우선 선택하였다. 이러한 처리를 통해 모델이 두 클래스를 균형있게 학습할 수 있도록 하였으며, 특히 소수 클래스인 불꽃에 대한 과적합을 방지하고자 하였다.

2. YOLOv8 모델 선택 및 학습 설정

1) YOLOv8 아키텍처 개요

본 연구에서는 Figure 4와 같이 Ultralytics에서 2023년 발표한 YOLOv8(You Only Look Once version 8) 시리즈 중 Large 모델(YOLOv8L)을 산불 탐지 모델의 기반 아키텍처로 선택하였으며, YOLOv8 모델의 특징은 다음과 같다.

(1) Feature Pyramid로 다양한 크기의 객체를 탐지

이미지는 CNN 기반 Backbone을 통해 여러 단계의 특징 맵(feature map)으로 변환되고, FPN·PAN 구조를 통해 작은 객체부터 큰 객체까지 모두 안정적으로 탐지한다.

(2) Anchor-free 방식

이전 YOLOv5/YOLOv7에서는 anchor box 기반이었지만, YOLOv8은 anchor-free 탐지 방식을 채택했다. 이를 통해 다음과 같은 장점이 생긴다.

- 학습 속도 증가
- 하이퍼파라미터 감소 (anchor 튜닝 불필요)
- 작은 객체에 대한 일반화 성능 향상

(3) Decoupled Head (분리된 Classification / Regression)

YOLOv8의 Head는 다음 두 가지를 분리하여 처리한다.

- Classification Head : 어떤 객체인지 예측
- Regression Head : 중심점, 크기 등을 기반으로 bounding box 좌표 예측

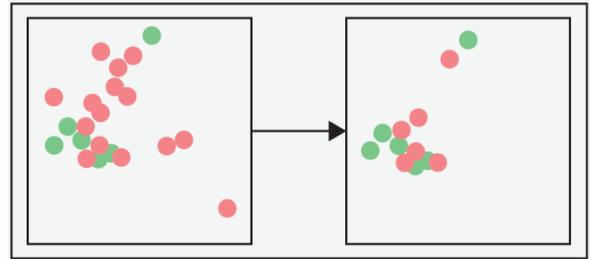


Figure 2. Downsampling.

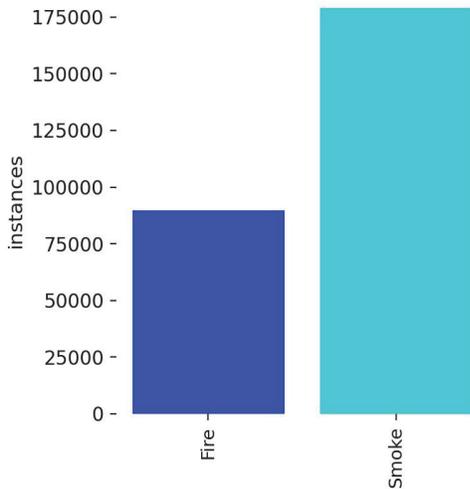


Figure 1. Class imbalance distribution.

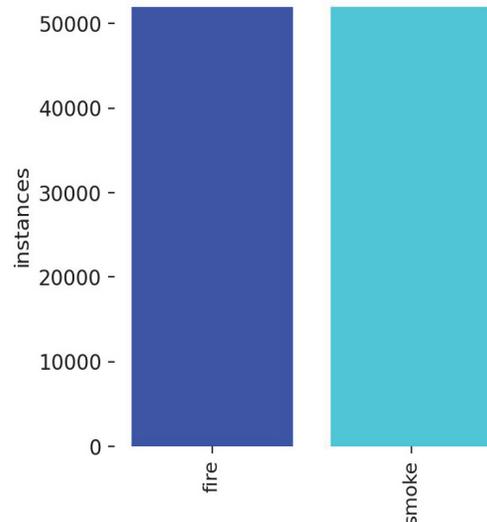


Figure 3. Balanced classes.

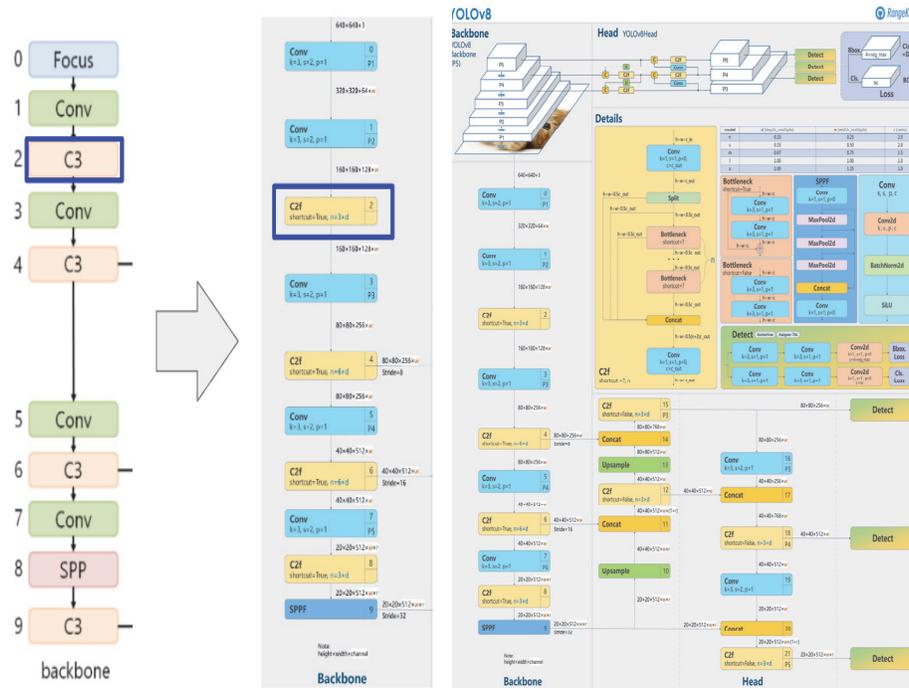


Figure 4. YOLOv8 architecture.

2) 주요 객체 탐지 모델 성능 비교 (COCO val2017 기준) 산불 조기 탐지 시스템은 초기 화재의 미약한 불꽃이나 희미한 연기를 정확하게 탐지해야 하며, 동시에 실시간 모니터링을 위한 빠른 처리 속도가 요구된다. 본 연구에서는 정확도와 속도의 균형을 고려하여 주요 객체 탐지 모델들과 비교 분석을 수행하였다.

Table 4와 같이 RT-DETR-L은 mAP50 53.0%로 가장 높은 정확도를 보였으나 추론 시간이 29.4 ms로 느리고 GFLOPs가 92.0으로 낮아 복잡한 환경 처리에 한계가 있다. Faster R-CNN은 2-stage 방식으로 mAP50 42.0%의 준수한 정확도를 보였으나, 추론 시간이 83.3 ms로 실시간 모니터링에 부적합하다. YOLOv6과 YOLOv7은 빠른 추론 속도를 제공하지만 mAP50이 각각 49.0%, 51.2%로 상대적으로 낮다. YOLOv8은 추론 시간 43.7 ms, mAP50 52.9%, GFLOPs 165.2의 균형잡힌 성능을 제공한다. RT-DETR-L 대비 1.9%p 낮은 정확도를 보이지만 약 1.5배 빠른 추론 속

도와 1.8배 높은 계산량을 통해 복잡한 화재·연기 패턴을 효과적으로 학습할 수 있다. YOLO 시리즈 중 가장 높은 mAP50(52.9%)를 기록하면서도 실시간 처리가 가능한 속도(약 22.9 FPS)를 달성하였다. 산불 감시 시스템은 넓은 지역을 장시간 모니터링해야 하므로 높은 정확도와 빠른 추론 속도가 필수적이다. 이러한 이유로 본 연구에서는 정확도, 추론 속도, 계산 효율성의 최적 균형점을 제공하는 YOLOv8i 모델을 최종 선택하였다.

### 3) 학습 파라미터 설정

Table 5와 같이 학습 에포크는 학습 시간 및 컴퓨팅 리소스 제약을 고려하여 200으로 설정하였으며, 배치 크기는 64 이상으로 설정 시 Out of Memory(OOM) 오류가 발생하였기 때문에 24GB VRAM의 메모리 한계를 고려하여 32로 설정하였다. 옵티마이저는 YOLOv8의 auto 설정에 따라 SGD(Stochastic Gradient Descent)가 자동으로 선택되었다.

Table 4. Model performance comparison.

Model	Type	Parameters (M)	GFLOPs	mAP50(%)	Inference Time (ms)
YOLOv5l	1-stage	46.5	109.1	49.0	6.7
YOLOv7	1-stage	36.9	104.7	51.2	6.2
<b>YOLOv8l</b>	<b>1-stage</b>	<b>43.7</b>	<b>165.2</b>	<b>52.9</b>	<b>8.4</b>
RT-DETR-L	Transformer	32.0	92.0	53.0	29.4
Faster R-CNN	2-stage	41.8	180.0	42.0	83.3

**Table 5. Hyperparameter Settings and Rationale.**

Parameter	Value	Rationale & Effect
epochs	200	Sufficient convergence; performance improvement observed even after 150 epochs
batch	32	Balance between GPU memory (24GB) efficiency and training stability (OOM prevention)
optimizer	Auto(SGD)	Momentum 0.937, automatically selected for batch size 32
lr0	0.01	Initial learning rate (YOLO standard)
lrf	0.01	Final learning rate ratio (1% of initial rate)
weight_decay	0.0005	Prevents overfitting via L2 regularization

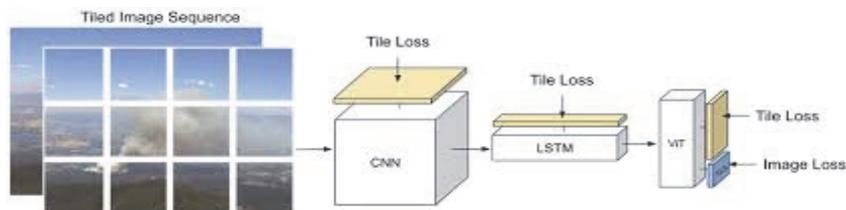
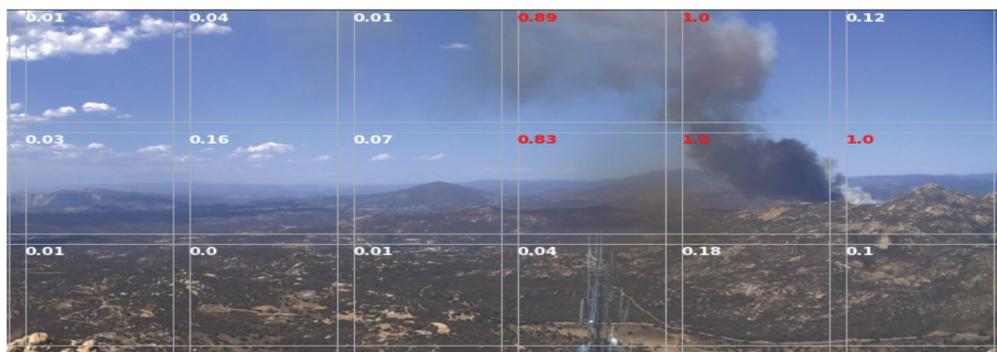
YOLOv8은 배치 크기가 64 미만인 경우 SGD를, 64 이상인 경우 AdamW를 사용하도록 설계되어 있다. 본 연구에서는 배치 크기 32로 인해 SGD with Momentum(0.937)이 적용되었다. 학습률은 초기 0.01에서 최종 0.0001로 선형 감소하도록 설정하였으며, 에포크 0-3 동안 워업 단계를 거쳐 사전학습 가중치의 안정화를 도모하였다. Weight decay는 0.0005로 설정하여 과적합을 방지하였다. COCO 데이터셋으로 사전학습된 YOLOv8l 가중치(yolov8l.pt)를 활용하여 전이학습을 수행하였으며, 검증 데이터셋에서 최고 mAP50을 기록한 모델을 최종 모델로 선택하였다. 최종 모델은 총 112개의 레이어와 43,607,584개의 파라미터로 구성되며, 연산량은 164.8 GFLOPs이다.

### 3. 2단계 검출 방식 설계

연기 탐지의 경우 높은 정확도 달성 이후에도 실제 화재에서 발생한 연기인지, 구름이나 안개 등의 자연 현상인지 구분

하기 어려운 한계가 예상되었다. 특히 단일 프레임 기반 탐지로는 정적인 구름과 동적으로 움직이는 연기를 구별하기 어렵고, 연기의 투명성, 희미함, 비정형적 형태 등으로 인해 오탐(False Positive)이 발생할 가능성이 있다. 이러한 문제를 해결하기 위해 YOLO 모델로 1차 검출된 연기 영역에 대해 2차 분석 단계를 추가하였다. 2차 분석에는 산불 연기 탐지에 특화된 SmokeyNet 모델을 적용하였다. Figure 5는 SmokeyNet의 전체 아키텍처를 보여준다. SmokeyNet은 고정 카메라 이미지에서 실시간 산불 연기를 감지하기 위해 개발된 딥러닝 아키텍처로, CNN(Convolutional Neural Network), LSTM(Long Short-Term Memory), 그리고 ViT(Vision Transformer)를 결합한 시공간 정보 처리 구조를 갖는다.

이 모델은 이미지를 224×224 크기의 타일로 분할하여 처리하며, 연속된 두 프레임의 시간적 정보를 LSTM으로 학습하고, ViT를 통해 타일 간 공간적 관계를 파악한다. SmokeyNet

**Figure 5. SmokeyNet architecture.****Figure 6. SmokeyNet results.**

은 약 25,000장의 산불 연기 이미지로 구성된 FigLib 데이터셋에서 학습되어 높은 정확도와 빠른 화재 감지 성능을 보고한 바 있다. 특히 투명하고 비정형적인 연기의 특성, 희미하거나 작은 연기 기둥, 그리고 구름이나 안개 등의 오탐 요인을 구별하는 데 강점을 보인다. Figure 6은 SmokeyNet의 연기 탐지 결과 예시를 보여준다.

본 연구에서는 YOLO 기반 1차 탐지 결과를 SmokeyNet으로 재검증함으로써 오탐률을 낮추고, 연기의 시간적 변화 패턴과 공간적 분포 특성을 종합적으로 분석하여 실제 화재 연기 여부를 보다 정확히 판별하고자 하였다. SmokeyNet은 연속된 프레임의 시공간 정보를 활용하여 정적인 구름/안개와 동적인 연기를 구분하는 데 강점을 가지고 있어, 단일 프레임 기반 YOLO의 한계를 보완할 수 있다. 이러한 2단계 검출 방식은 초기 화재 탐지의 신뢰성을 높이고 오탐으로 인한 불필요한 경보를 줄이며, 화재 대응 시간을 단축하는 데 기여할 것으로 기대된다.

## 연구 결과

### 1. 초기 모델 학습 결과

초기 모델 학습 수행 결과, 불꽃 탐지 평균정확도(mAP50)는 0.977, 연기 탐지 평균정확도는 0.684, 전체 객체 탐지 평균정확도는 0.830으로 나타났다. 여기서, mAP50은 두 바운딩박스가 겹치는 비율(IoU) 50% 이상인 예측값 대비 실제 검출(True positive)된 클래스의 평균정확도이다. 불꽃 탐지 정확도가 상대적으로 높게 나타난 것은 학습데이터의 수량이 적어 모델 학습 시 과적합(overfitting)이 주요 원인으로 추정되었다. 다시 말해서, 탐지 대상 이미지 내 불꽃 위치 탐지는 평균 대비 높은 인식이 된 반면, 연기 탐지 정확도는 상대적으로 낮게 나타났다.

### 2. 클래스 균형 조정을 통한 성능 개선

클래스 불균형 문제 해결을 위해 다운샘플링 기법을 적용한 후 모델을 재학습하였다. 최종 모델은 112개의 레이어와 약 4,360만 개의 파라미터로 구성되었으며, 연산량은 164.8 GFLOPs로 측정되었다. 총 34,127장의 검증 이미지와 46,284개의 객체 인스턴스(불꽃 13,786개, 연기 32,498개)를 대상으로 성능을 평가한 결과를 Table 6에 제시하였다. 전체 평균정확도

(mAP50)는 0.756로 나타났다. 클래스별로 분석하면, 불꽃 탐지는 Precision 0.836, Recall 0.748, mAP50 0.796를 기록하였고, 연기 탐지는 Precision 0.818, Recall 0.626, mAP50 0.717로 측정되었다.

다운샘플링 기법의 적용으로 초기 모델 대비 과적합 현상이 크게 완화되었으며, 불꽃과 연기 클래스 간의 성능 격차가 감소하였다. 특히 불꽃 탐지의 경우 높은 재현율(Recall 0.748)을 보여 실제 화재 발생 시 놓치는 경우가 적음을 확인하였다. 연기 탐지는 정밀도(Precision 0.818)가 높게 나타나 오탐률이 상대적으로 낮았으나, 재현율(Recall 0.626)이 다소 낮아 희미한 연기나 먼 거리의 연기를 놓치는 경우가 있었다.

다만, 다운샘플링 기법의 적용으로 성능이 극적으로 증가하지는 않았으나, 학습 과정에서 손실 함수의 변동성이 감소하고 검증 데이터에 대한 일반화 성능이 개선되어 전반적으로 안정적인 학습 양상을 보였다. 이는 클래스 불균형으로 인한 편향이 완화되면서 모델이 소수 클래스와 다수 클래스를 균형있게 학습할 수 있게 되었음을 의미한다. 향후 AI허브 및 추가 데이터 소스를 통해 양질의 학습 데이터가 더욱 확보된다면, 균형잡힌 데이터셋 구성을 바탕으로 성능이 더욱 향상될 것으로 판단된다.

Figure 7, 8, 9, 10은 최종 YOLOv8l 모델의 학습 결과를 나타낸다. 4개 그래프는 신뢰도 임계값(Confidence Threshold)에 따른 성능 지표 변화를 보여주며, 하단의 혼동 행렬은 클래스별 분류 정확도를 시각화한다. F1-Confidence Curve는 F1 Score가 신뢰도 임계값 0.337에서 최댓값 0.77을 기록함을 보여준다. 이는 모델이 약 34%의 확신을 가질 때 정밀도와 재현율이 가장 균형잡힌 성능을 발휘함을 의미한다. Precision-Confidence Curve에서는 임계값이 증가할수록 정밀도가 상승하여 0.701에서 약 0.95의 최댓값을 달성하였으며, 이는 높은 확신도를 가진 예측일수록 정확도가 높음을 나타낸다.

반면 Recall-Confidence Curve는 임계값 증가 시 재현율이 급격히 감소하는 전형적인 Trade-off 관계를 보여주며, 낮은 임계값에서 최대 0.82의 재현율을 기록하였다. Precision-Recall Curve는 정밀도와 재현율의 종합적인 관계를 나타내며, mAP@0.5는 0.806을 달성하였다. 곡선이 우상단에 위치할수록 우수한 성능을 의미하며, 본 모델은 대부분의 구간에서 높은 정밀도를 유지하면서도 준수한 재현율을 보였다.

Table 6. Final performance metrics after class balancing.

Class	Images	Instances	Precision	Recall	mAP50
Overall	34,127	46,284	0.827	0.687	0.756
Fire	10,564	13,786	0.836	0.748	0.796
Smoke	26,334	32,498	0.818	0.626	0.717

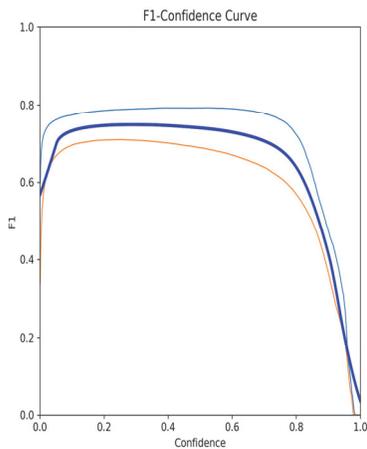


Figure 7. F1 Confidence\_Curve.

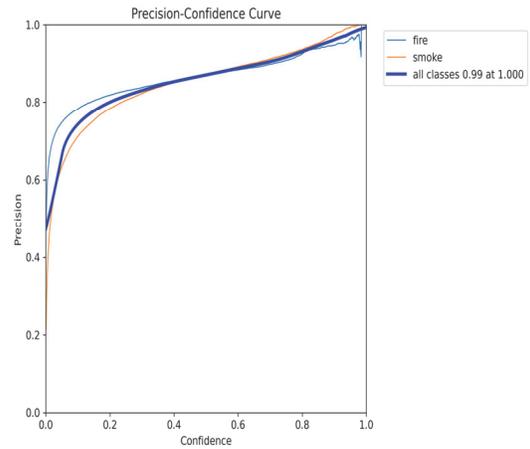


Figure 8. Precision Confidence\_Curve.

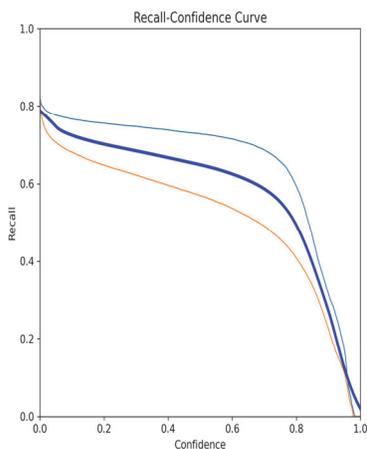


Figure 9. Recall Confidence\_Curve.

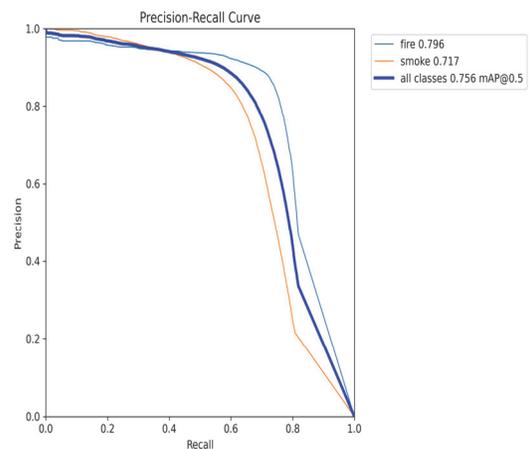


Figure 10. Precision-Recall Confidence\_Curve.

Figure 11의 정규화된 혼동 행렬(Normalized Confusion Matrix)은 클래스별 분류 성능을 정량적으로 보여준다. 불꽃(Fire) 클래스는 75%의 정확도로 올바르게 분류되었으나, 24%는 배경(Background)으로 오분류되었다. 연기(Smoke) 클래스는 72%의 정확도를 보였으며, 26%가 배경으로 잘못 분류되었다. 이러한 결과는 모델이 배경과 화재·연기를 구분하는 능력은 우수하나(Specificity 94%), 희미한 화재나 먼거리의 연기를 배경으로 오인하는 경우가 있음을 나타낸다. 특히 불꽃과 연기가 배경으로 오분류되는 비율(24%, 26%)이 재현율 저하의 주요 원인으로 분석된다.

### 3. 시스템 통합 성능 평가

본 연구에서는 제안한 2단계 화재 탐지 시스템의 성능을 정량적으로 평가하기 위해 Precision, Recall, F1 Score, Accuracy의 평가 지표에 Specificity를 추가하여 활용하였다. YOLO 단독 탐지의 경우 1920×1080 해상도 환경에서 평균

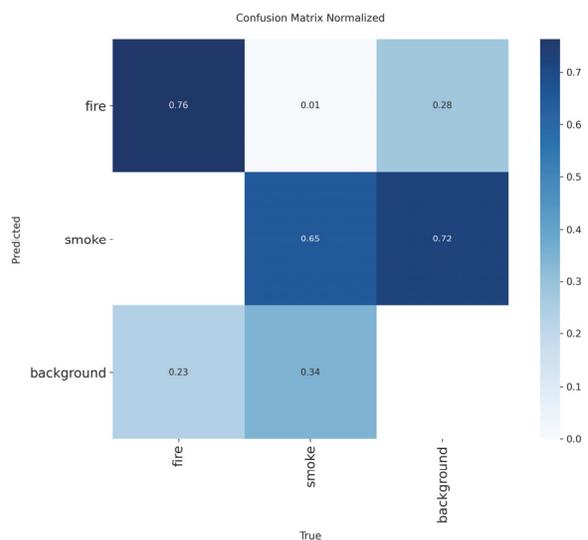


Figure 11. Normalized confusion matrix.

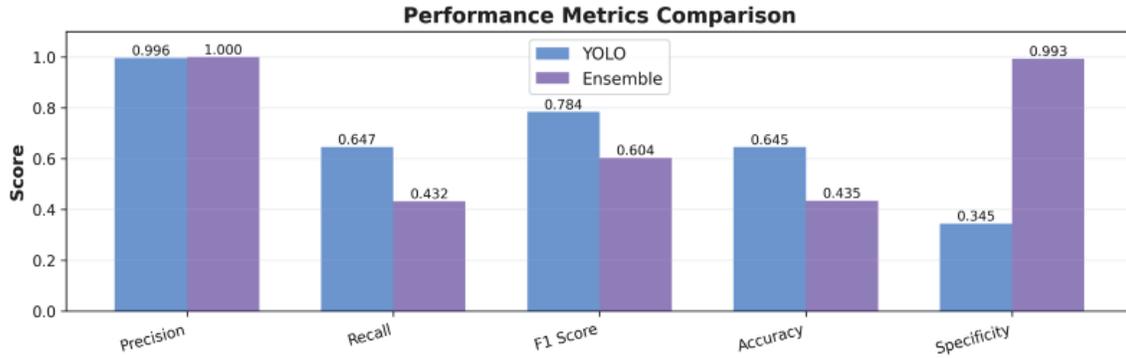


Figure 12. Evaluation metric scores.

95.06 FPS를 기록하며 실시간 처리 성능을 확인하였다. 반면, YOLO+SmokeyNet 앙상블 모델은 추가 검증 과정으로 인해 처리 속도가 저하되었지만, 탐지 신뢰도 측면에서 유의미한 개선을 달성하였다.

2단계 검출 시스템의 성능을 검증하기 위해 34,132개 샘플 (총 46,284개 객체 인스턴스)로 구성된 대규모 데이터셋을 사용하여 YOLO 단독 모델과 YOLO+SmokeyNet 앙상블 모델을 비교 평가하였다. Figure 12는 각 모델의 평가지표 점수를 비교한 결과이다. YOLO 단독 모델은 재현율(Detection Rate) 64.67%, 정밀도 99.58%, 특이도(Specificity) 34.51%를 기록하였다. 높은 재현율과 정밀도에도 불구하고 낮은 특이도(34.51%)는 실제 화재가 아닌 경우를 화재로 오판하는 비율이 높음을 의미하며, 이는 93건의 오탐(False Positive)으

로 나타났다. 반면 YOLO+SmokeyNet 앙상블 모델은 재현율 43.23%, 정밀도 99.99%, 특이도 99.30%를 달성하였다. 재현율은 21.44%p 감소하였으나, 특이도는 64.79%p 향상되어 오탐이 단 1건으로 감소하였다. 이는 SmokeyNet의 2차 검증 단계가 YOLO의 오탐을 효과적으로 필터링하여 초기 화재 탐지의 신뢰성을 크게 향상시켰음을 의미한다.

2단계 검출 방식 성능 비교

YOLO 단독 모델과 YOLO+SmokeyNet 앙상블 모델의 성능을 정량적으로 비교한 결과는 Table 7과 같다.

Table 7에서 보는 바와 같이, YOLO 단독 모델의 Detection Rate는 64.70%였으나 평균 신뢰도는 34.70%로 낮아 오탐(False Positive)이 빈번하게 발생하였다. 반면,

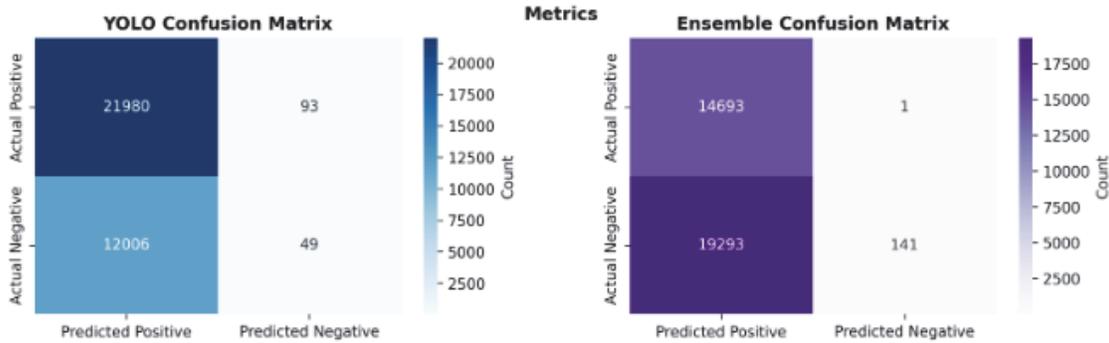


Figure 13. Confusion Matrix Comparison: YOLOv8l vs. YOLOv8l+SmokeyNet.

Table 7. YOLO-only vs YOLO + SmokeyNet Ensemble Performance.

Metric	YOLO Only	YOLO + SmokeyNet	Change
Detection Rate	64.70%	43.2%	-21.5%
Avg Confidence	34.70%	88.18%	+53.49%
Inference Time	10.52ms	466.58ms	+456.06ms
FPS	95.06	2.14	-

YOLO+SmokeyNet 앙상블 모델은 Detection Rate가 43.2%로 감소하였지만, 평균 신뢰도가 88.18%로 크게 향상되었다.

Figure 13은 YOLO 단독 모델과 YOLO+SmokeyNet 앙상블 모델의 혼동 행렬을 비교한다. YOLO 단독 모델은 21,980건의 True Positive(실제 화재를 정확히 탐지)와 93건의 False Positive(오탐)를 기록하였으나, False Negative(미탐지)가 12,006건으로 높고 True Negative는 49건에 불과하였다. 이는 YOLO가 높은 재현율(64.67%)을 달성했지만, 실제 화재가 아닌 경우를 제대로 구분하지 못해 특이도가 34.51%에 그쳤음을 의미한다. 반면 YOLO+SmokeyNet 앙상블 모델은 True Positive가 14,693건으로 감소하였으나, False Positive는 93건에서 단 1건으로 급격히 감소하였고, True Negative는 49건에서 141건으로 약 2.9배 증가하였다. 이는 2차 검증 단계인 SmokeyNet이 YOLO의 오탐 93건 중 92건(98.9%)을 효과적으로 제거하고, 실제 화재가 아닌 경우를 정확히 구분하는 능력(특이도 99.30%)을 크게 향상시켰음을 보여준다.

#### 하드웨어 · 데이터 품질 한계에 따른 영향 반영

현재 하드웨어 성능의 한계로 인해 Epoch, Batch size 등 주요 하이퍼파라미터를 충분히 설정하지 못하고 있으며, 일부 데이터 품질 또한 낮아 전체적인 탐지 성능이 제한되고 있다. 이 제한들로 인해 탐지율(Recall)은 잠재적 상한선에 도달하지 못한 상태이며, 향후 하드웨어 성능을 개선하고 데이터 품질을 향상시켜 탐지율을 일정 수준 이상으로 안정화한다면, 그 이후에는 오탐(False Positive)이 모델 품질을 결정하는 핵심 요인이 된다(Smith et al., 2018; Trigka and Dritsas, 2025). 특히 화재 · 연기와 같은 희귀 이벤트 탐지 분야에서는 오탐률이 시스템의 실사용 가능성을 좌우하는 가장 중요한 지표로 보고된다. 따라서 탐지율이 개선된 이후에는 신뢰도 향상과 오탐률 감소가 모델 고도화의 핵심 방향으로 요구된다(Lee and Shim, 2019; Joshi et al., 2025).

#### 모델 파라미터 설정

본 연구에서 사용한 주요 파라미터 설정은 Table 8과 같다.

YOLO 모델의 신뢰도 임계값은 0.25로 설정하여 초기 검출 단계에서 민감하게 반응하도록 하였으며, SmokeyNet의 확률 임계값은 0.5로 설정하여 2차 검증 단계에서 엄격하게 필터링하도록 구성하였다. 앙상블 가중치는 다양한 실험을 통해 YOLO 0.6, SmokeyNet 0.4로 설정하였을 때 최적의 성능을 나타냄을 확인하였다. SmokeyNet의 타일 그리드는 FlgLib 데이터셋 논문에서 제시한 5×5 구조를 따랐으며, 타일 크기는 ImageNet 표준인 224×224 픽셀로 설정하였다.

#### 종합 성능 평가

테스트 데이터셋 500장을 대상으로 평가한 결과, Precision은 66.5%, Recall은 61.9%, F1-score는 64.1%로 나타났다. False Positive Rate는 33.5%로 나타나 오탐 감소를 위한 추가적인 개선이 필요함을 확인하였다. 2단계 검출 방식의 도입으로 탐지 신뢰도는 크게 향상되었으나, 처리 속도 저하가 실시간 모니터링 시스템 적용에 제약 요인으로 작용할 수 있어, 향후 모델 경량화 및 하드웨어 가속 최적화 연구가 필요할 것으로 판단된다.

## 결론

본 연구에서는 초기 화재 탐지의 정확도를 높이기 위해 딥러닝 기반 불꽃 및 연기 탐지 모델을 개발하고, 클래스 불균형 문제 해결과 오탐 감소를 위한 2단계 검출 방식을 제안하였다.

초기 YOLO 모델 학습 결과, 불꽃 탐지는 mAP50 0.977의 높은 정확도를 보인 반면, 연기 탐지는 0.684로 상대적으로 낮은 성능을 나타냈다. 이는 학습데이터의 클래스 불균형으로 인한 과적합이 주요 원인으로 분석되었다. 이를 해결하기 위해 AI허브 공공데이터셋을 활용하여 불꽃 데이터를 추가

Table 8. Model parameter configuration.

Category	Parameter	Value	Source / Rationale
YOLO	Confidence Threshold	0.25	YOLO default
SmokeyNet	Probability Threshold	0.5	Binary classification standard
Ensemble	YOLO Weight	0.6	Experimentally optimized
Ensemble	SmokeyNet Weight	0.4	Experimentally optimized
Ensemble	Detection Threshold	0.5	Binary classification standard
SmokeyNet	Tile Grid	5×9	FlgLib paper setting
SmokeyNet	Tile Size	224×224	ImageNet standard

Table 9. Single model vs ensemble model comparison.

Model	Precision	Recall	F1 Score	Accuracy	Specificity
YOLO	99.6%	64.7%	78.4%	64.5%	34.5%
YOLO+SmokeyNet	99.9%	43.2%	60.4%	43.5%	99.3%

수집하고, 다운샘플링 기법을 통해 다수 클래스인 연기 데이터를 소수 클래스인 불꽃 데이터 수준으로 조정하여 클래스 간 균형을 맞추어 학습한 결과 안정성이 크게 개선된 것을 확인하였다. 그러나 단일 프레임 기반 YOLO 탐지만으로는 실제 화재 연기와 구름, 안개 등 자연 현상을 구분하기 어려운 한계가 있었다. 이에 본 연구에서는 YOLO 기반 1차 탐지와 SmokeyNet 기반 2차 검증을 결합한 2단계 검증 방식을 제안하였다. SmokeyNet은 CNN, LSTM, ViT를 결합한 시공간 정보 처리 구조를 통해 연속된 프레임의 시간적 변화 패턴과 공간적 분포 특성을 종합적으로 분석함으로써, 정적인 구름·안개와 동적인 연기를 효과적으로 구별할 수 있다.

2단계 검출 시스템의 성능을 검증하기 위해 34,132개 샘플(총 46,284개 객체 인스턴스)로 구성된 대규모 데이터셋을 사용하여 YOLO 단독 모델과 YOLO+SmokeyNet 앙상블 모델을 비교 평가하였다.

YOLO 단독 모델은 재현율(Detection Rate) 64.67%, 정밀도 99.58%, 특이도(Specificity) 34.51%를 기록하였다. 높은 재현율과 정밀도에도 불구하고 낮은 특이도(34.51%)는 실제 화재가 아닌 경우를 화재로 오판하는 비율이 높음을 의미하며, 이는 93건의 오탐(False Positive)으로 나타났다.

반면 YOLO+SmokeyNet 앙상블 모델은 재현율 43.23%, 정밀도 99.99%, 특이도 99.30%를 달성하였다. 재현율은 21.44%p 감소하였으나, 특이도는 64.79%p 향상되어 오탐이 단 1건으로 감소하였다. 이는 SmokeyNet의 2차 검증 단계가 YOLO의 오탐을 효과적으로 필터링하여 초기 화재 탐지의 신뢰성을 크게 향상시켰음을 의미한다(Table 9).

본 연구의 주요 기여는 다음과 같다. 첫째, 클래스 불균형 문제를 데이터 샘플링 기법을 통해 효과적으로 해결하여 YOLO 모델의 안정성과 일반화 성능을 개선하였다. 둘째, 단일 모델의 한계를 보완하기 위한 2단계 검출 방식을 제안하여 오탐률을 93건에서 1건으로 감소시키고 특이도를 99%대로 향상시켰다. 셋째, 총 46,284개의 객체 인스턴스를 포함하는 대규모 데이터셋으로 검증된 실용적인 시스템 구조를 제시하였다.

화재·연기와 같은 희귀 이벤트(rare event) 탐지 분야에서는 오탐률(False Positive Rate)이 실사용 가능성을 결정짓는 핵심 요인이 여러 선행연구에서 강조된다. 본 연구의 앙상블 모델은 오탐을 99% 이상 감소시켰으로써 실제 화재 감시 시스템에 적용 가능한 수준의 신뢰성을 확보하였다. 재현율

의 감소(64.67% → 43.23%)는 보수적 탐지 전략의 결과이며, 향후 더 높은 Batch size, 더 긴 학습 Epoch, 더 우수한 데이터 품질 확보를 통해 재현율과 특이도를 동시에 향상시킬 수 있을 것으로 기대된다.

다만, 본 연구는 몇 가지 한계점을 가지고 있다. 첫째, 현재 학습 데이터의 양적 제약으로 인해 다운샘플링 기법의 효과가 제한적이었으며, 특히 연기 탐지의 재현율(0.626)이 불꽃 탐지(0.748)에 비해 낮아 희미한 연기나 먼 거리의 연기를 놓치는 경우가 있었다. 둘째, SmokeyNet 적용에 따른 추론 시간이 573.55 ms로 증가하여 실시간 처리 속도가 1.74 FPS로 저하되었으므로, YOLO 단독 모델의 46.17 FPS 대비 약 26.5배 느린 속도는 실시간 모니터링 시스템 적용 시 병목이 될 수 있다. 셋째, 야간이나 저조도 환경, 극심한 기상 조건에서의 탐지 성능에 대한 추가 검증이 필요하며, 다양한 현장 환경에서의 실증 테스트가 요구된다.

## 제 언

향후 연구에서는 다음과 같은 방향으로 시스템을 개선할 계획이다. 첫째, 더욱 다양한 환경 조건에서의 데이터셋 확보와 데이터 증강 기법을 통해 연기 탐지의 재현율을 향상시킬 것이다. 둘째, 양질의 데이터와 성능 좋은 하드웨어를 확보하여 recall 지표를 상승 시키고, 앙상블 시스템의 추론 속도를 향상시켜, GPU 가속 및 병렬 처리 최적화를 통해 실시간 처리가 가능한 수준으로 발전시킬 것이다. 셋째, 드론 기반 이동형 화재 탐지 시스템으로의 확장 가능성을 탐색하고, IoT 센서 데이터(온도, 습도, 풍향)와의 융합을 통한 다중 모달 화재 탐지 시스템을 개발할 것이다. 본 연구가 제시한 2단계 검출 방식은 화재 조기 경보 시스템의 오탐률을 획기적으로 감소시켰으므로 시스템 신뢰성을 크게 향상시켰다. 이는 산불 피해를 최소화하고 신속한 대응 체계 구축에 기여할 것으로 기대되며, 실제 산불 감시 현장에서 운용 가능한 실용적인 솔루션으로 발전할 수 있는 기반을 마련하였다.

## 감사의 글

본 연구는 산림청(한국임업진흥원) ‘산림과학기술 연구개발사업(RS-2024-00402624)’의 지원에 의하여 이루어진 것입니다.

## References

- Al-Smadi, Y., Alauthman, M., Al-Qerem, A., Aldweesh, A., Quaddoura, R., Aburub, F., Mansour, K. and Alhmiedat, T. 2023. Early Wildfire Smoke Detection Using Different YOLO Models. *Machines* 2023, 11: 246.
- Dewangan, A., Pande, Y., Braun, H.W., Vernon, F., Perez, I., Altintas, I., Cottrell, G.W. and Nguyen, M.H. 2022. FlgLib & SmokeyNet: Dataset and Deep Learning Model for Real-Time Wildland Fire Smoke Detection. *Anshuman Dewangan, Remote Sensing* 14: 1007.
- He, L., Zhou, Y., Liu, L., Zhang, Y. and Ma, J. 2025. Research and application of deep learning object detection methods for forest fire smoke recognition. *Scientific Reports* 15: 16328.
- Joshi, A.S., William, G.J. and Joshi, S.S. 2025. Uncertainty-Aware Post-Detection Framework for Enhanced Fire and Smoke Detection in Compact Deep Learning Models. *International Conference on Smart Multimedia arXiv:2510.10108*
- Lee, Y. and Shim, J. 2019, False Positive Decremental Research for Fire and Smoke Detection in Surveillance Camera using Spatial and Temporal Features Based on Deep Learning. *Electronics* 8: 1167.
- Mahdi, A. S. and Mahmood, S. A. 2022, Analysis of Deep Learning Methods for Early Wildfire Detection Systems: Review. 2022 5th International Conference on Engineering Technology and its Applications (IICETA), Al-Najaf, Iraq, 271-276.
- Maria Trigka, Elias Dritsas, 2025 “A Comprehensive Survey of Deep Learning Approaches in Image Processing”, *Sensors*, 25, 531.
- Park, J.W., Yoon, H.J., Lee, B.D., Woo, C.S, Kim, Y.B. and Lee, B.N, 2016. Review and Recommendations on Wildfire Smoke Dispersion Models for the Development of a Wildfire Smoke Impact Forecasting System in Korea. *Journal of the Korean Society of Hazard Mitigation* 16: 197-208.
- Smith, S. L., Kindermans, P.J., Ying, C. and Le, Q.V. 2018. Don't Decay the Learning Rate, Increase the Batch Size. *The International Conference on Learning Representations*, arXiv:1711.00489v2
- Wang, X., Li, M., Gao, M., Liu, Q., Li, Z. and Kou, L. 2023. Early smoke and flame detection based on transformer. *Journal of Safety Science and Resilience* 4(3): 294-304.
- Wasserman, T. N. and Mueller, S.E. 2023. Climate influences on future fire severity: a synthesis of climate-fire interactions and impacts on fire regimes, high-severity fire, and forests in the western United States. *Fire Ecology* 19: 43
- Zhao, J., Zheng, B., Ciais, P., Chen, Y., Gasser, T., Canadell, J.G., Zhang, L. and Zhang, Q. 2025. Global warming amplifies wildfire health burden and reshapes inequality. *Nature* 647: 928-934.

---

Manuscript Received : November 6, 2025

First Revision : December 3, 2025

Second Revision : December 11, 2025

Accepted : December 11, 2025