

Design and Performance Analysis of a Smoke-Specialized RT-DETR Improved Model for Real-Time Wildfire Monitoring Using Drone Imagery

Young-Bok Cho*

*Professor, Dept. of Computer Education, Gyeongkok National University, Andong, Korea

[Abstract]

Early detection of forest fire smoke using drone imagery is essential for real-time monitoring of vast forested areas, but applying general object detectors faces limitations due to drone-specific challenges such as altitude variations, viewpoint diversity, and small-scale smoke objects. In this paper, we propose an improved model with a smoke-specialized backbone based on RT-DETR for real-time forest fire monitoring using drone imagery, accompanied by rigorous performance analysis. The proposed model configures its backbone with SCPCConvBlock, which combines the channel-significance-based selective convolution module (SCPCConv) and Efficient Multi-Scale Attention (EMA), and applies a foreground-focused multi-scale feature pyramid network (MSFFPN) to resolve the background confusion problem of small and sparse smoke frequently occurring in drone footage. Comprehensive comparative and ablation experimental results on a drone-captured forest fire smoke dataset demonstrate that the proposed model achieved superior performance with Precision of 0.883 (+0.7%p), Recall of 0.800 (+2.4%p), mAP50 of 0.862 (+3.8%p), and mAP95 of 0.539 (+2.2%p) compared to the baseline RT-DETR, while significantly accelerating training convergence speed by approximately 80% (from epoch 125 down to 25). We explicitly state as a limitation that these results are derived from a single execution under fixed hyperparameters and random seeds, and statistical variability across multiple runs was not evaluated.

► **Key words:** Wildfire Smoke Detection, Drone Imagery, RT-DETR, SCPCConv, MSFFPN, Real-time Detection

[요 약]

드론 영상을 활용한 산불 연기의 조기 감지는 광범위한 산림 지역을 실시간으로 감시하는 데 필수적이거나, 드론 특유의 고도 변화, 시점 다양성, 소형 연기 객체 등의 문제로 일반 객체 검출기 적용에 한계가 있다. 본 논문에서는 드론 영상 기반 실시간 산불 감시를 위해 RT-DETR을 기반으로 연기 특화 백본을 개선한 모델을 제안하고 그 성능을 면밀히 분석한다. 제안 모델은 채널 중요도 기반 선택적 합성곱 모듈(SCPCConv)과 효율적 다중 스케일 어텐션(EMA)을 결합한 SCPCConvBlock을 백본으로 구성하고, 전경 집중 다중 스케일 피라미드 네트워크(MSFFPN)를 적용하여 드론 영상에서 빈번히 발생하는 소형·희박 연기의 배경 혼동 문제를 해소하였다. 드론 촬영 산불 연기 데이터셋에서의 비교 및 절제 실험 결과, 제안 모델은 RT-DETR 대비 정밀도(Precision) 0.883(+0.7%p), 재현율(Recall) 0.800(+2.4%p), mAP50 0.862(+3.8%p), mAP95 0.539(+2.2%p)의 향상된 성능을 달성하였으며, 학습 수렴 에포크를 기존 125 에포크에서 25 에포크로 약 80% 단축하였다. 본 결과는 고정된 하이퍼파라미터와 난수 시드 하에서 수행된 단일 실험 결과이며, 반복 실행에 따른 통계적 변동성은 평가하지 않았음을 한계점으로 명시한다.

► **주제어:** 산불 연기 검출, 드론 영상, RT-DETR, SCPCConv, MSFFPN, 실시간 감지

-
- First Author: Young-Bok Cho, Corresponding Author: Young-Bok Cho
 - Young-Bok Cho (ybcho@gknu.ac.kr), Dept. of Computer Education, Gyeongkok National University
 - Received: 2026. 06. 25, Revised: 2026. 07. 20, Accepted: 2026. 07. 22.

I. Introduction

산불이나 도시 화재와 같은 대형 화재는 인명과 재산뿐만 아니라 생태계에도 막대한 피해를 발생시키는데, 특히 산불은 초기 단계에서 빠르게 확산되고 진화하기 어렵기 때문에 연기 단계에서 가능한 한 빨리 화재를 감지하는 것이 매우 중요하다. 미국과 캐나다의 통계에 따르면 매년 수천 건의 사상자와 수백만 헥타르의 산림이 피해를 입는 것으로 보고되어 화재 감지 기술의 선제적 개발이 필요함을 나타내고 있다[1,2]. 산불의 사전 예측에 있어 현재 실내 환경에서는 광전 센서와 열 센서를 이용한 센서 기반 감지기가 주로 사용되지만, 산불과 같은 실외 환경에서는 연기와 열이 빠르게 희석되기 때문에 센서 기반 방식만으로는 충분한 민감도와 거리를 확보하기 어렵다. 이에 따라 CCTV, 드론, 위성 등 영상 데이터를 기반으로 한 비전 기반 연기 및 화재 감지 연구가 활발히 진행되고 있다[2,3]. 특히 드론은 기동성이 뛰어나고 광범위한 산림 지역을 신속하게 탐색할 수 있어 산불 조기 감지 플랫폼으로 주목받고 있으나, 드론 영상은 다양한 고도와 시점, 빠른 이동으로 인한 모션 블러, 소형 연기 객체 등 일반 카메라와는 다른 특성을 가지고 있어 전용 연기 검출 모델의 필요성이 제기된다[3,4]. 연기는(1) 흰색, 회색, 검은색과 같은 다양한 색상, (2) 불규칙한 공간적 형태, (3) 희미한 연기, 안개, 구름 등과 쉽게 혼동되는 특성을 가지고 있어, 기존의 특징 기반 알고리즘과 일반 객체 기반의 딥러닝 감지기에서는 높은 오탐지를 보이기 쉽다는 문제점이 있다[2,3,4]. 이러한 문제는 드론의 저고도 근접 촬영 시 배경과 연기의 대비가 낮아지는 상황에서 더욱 심화된다.

따라서 본 논문에서는 드론 영상의 특성을 고려한 연기 감지 트랜스포머 모델을 제안한다. 구체적으로 채널 중요도 기반SCPCConv, 효율적 다중 스케일 어텐션(EMA), 전경 집중 다중 스케일 피라미드(MSFFPN)를RT-DETR에 통합하여 드론 영상 기반 실시간 산불 연기 검출 성능을 향상시키는 것을 목표로 한다.

II. Preliminaries

1. Related works

1.1 Computer Vision and Deep Learning-Based Smoke Detection

컴퓨터 비전 기반 연기 검출에 관한 초기 연구들은 주로 색상, 텍스처, 움직임과 같은 수작업(hand-crafted) 특징

에 의존하여 영상 및 비디오에서 연기를 검출하는 방식으로 진행되었다. 정지 영상 기반 방법의 경우, 연기의 색상 계열, 텍스처, 공간 빈도 특성 등을 추출한 뒤SVM과 같은 고전적인 분류기를 이용해 연기 여부를 판별하거나, 비디오 기반 연구에서는 프레임 간 차분, 광류(optical flow) 등을 활용해 연기의 난류성 움직임을 특징으로 분류하는 연구가 지속되었다[3,4]. 이러한 전통적 접근법은 계산량이 적고 구현이 비교적 단순하다는 장점을 가지지만, 연기가 매우 얇거나 배경과의 명암 대비가 낮은 상황에서 오탐(false positive)과 누락(false negative)이 빈번하게 발생하는 문제점이 있다.

딥러닝 기법의 발전과 함께 연기 이미지를 대상으로 하는CNN 기반 분류기, 세멘틱 세그멘테이션(semantic segmentation) 모델, YOLO-Faster R-CNN 등 일반 객체 검출기를 활용한 연구들이 활발히 제안되었다. 그러나 이미지 분류 기반 방법은 연기의 공간적 위치를 실시간으로 추적·검출하는 데에 구조적인 제약이 따르며, 세그멘테이션 기반 방법은 연기 형태 자체가 모호하고 경계가 불분명하기 때문에 신뢰할 수 있는 라벨링이 어렵고 고해상도 영상 처리 시 계산 복잡도가 크게 증가한다는 문제가 있다[4,5]. 특히 드론 영상에서는 고도 변화에 따른 연기 크기 다양성과 시점 변화로 인해 기존 앵커 기반 검출기의 성능 저하가 두드러진다.

1.2 Transformer-based smoke detection

Transformer 기반 연기 검출 연구는 자연어 처리 분야에서 입증된 self-attention 메커니즘을 활용해 연기와 같이 비정형적이고 비국소적인 시공간 패턴을 갖는 대상의 표현력을 향상시키는 데 중요한 역할을 수행하고 있다[4,5]. Transformer의self-attention은 입력 전역에 걸쳐 상호 연관성을 직접 모형화함으로써 장거리 의존성 및 넓은 범위의 문맥 정보를 포착할 수 있어, 연기 패턴이 화면 전역으로 확산되거나 배경과 혼재되는 경우가 많은 드론 영상 환경에 특히 적합하다.

DETR(Detection TRansformer)는 Transformer를 이용하여 객체 검출을 End-to-End 방식으로 수행하는 대표적인 구조로, 사전에 정의된 앵커 박스와 비최대 억제(NMS) 과정 없이 object query를 통해 후보 객체를 직접 예측하는 방식을 취한다[5]. 이후 수렴 속도 개선과 다중 스케일 처리를 위해 Deformable DETR, DAB-DETR, DN-DETR, RT-DETR 등 다양한 파생 구조가 제안되었으며, RT-DETR은 실시간 성능을 목표로 하이브리드 인코더와 IoU-aware query selection을 도입하여 YOLO 계열에 준

하는 추론 속도와 우수한 검출 정확도를 달성하였다[5,7].

1.3 Drone Image-Based Wildfire Detection Research

드론을 활용한 산불 및 연기 감지 연구는 위성·CCTV 기반 연구와 달리 저고도 실시간 영상 처리의 관점에서 진행되고 있다. Lee et al.(2024)는 Swin Transformer를 기반으로 Planetscope와 Sentinel-2 위성 영상을 결합한 산불 피해 감지 연구를 수행하였으나, 위성 영상의 낮은 시간 해상도로 인해 실시간 초기 감지에는 제약이 있었다 [6]. 드론 기반 연구에서는 소형 연기 객체의 검출이 핵심 과제로 부각되며, 기존 YOLO 계열 검출기가 연기의 비정형 형태와 낮은 대비를 충분히 포착하지 못하는 한계가 보고되었다[3]. 이에 본 연구는 드론 영상의 특성을 반영한 연기 특화 백본 구조와 전경 집중 피라미드 네트워크를 통해 이러한 한계를 극복하고자 한다.

III. The Proposed Scheme

1. System Architecture

본 연구에서 제안하는 드론 영상 기반 산불 연기 검출 모델의 전체 구조는 <그림1>과 같다.

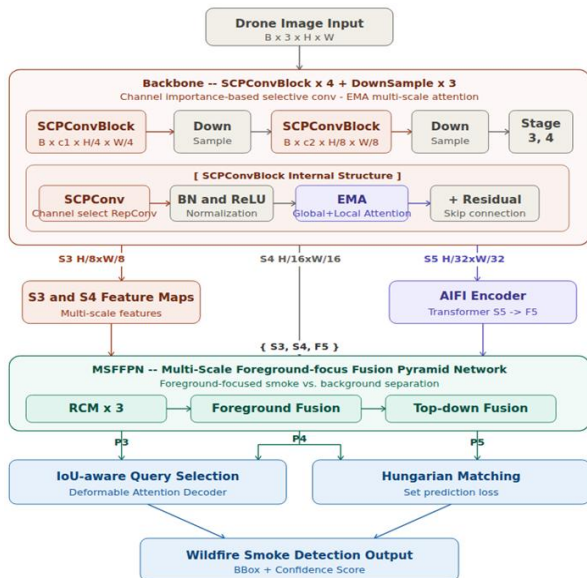


Fig. 1. Dron-based wildfire smoke detection transformer model backbone architecture

드론으로 촬영된 입력 영상은 먼저 백본 네트워크를 통과하면서 서로 다른 해상도를 갖는 세 가지 다중 스케일 특징 맵 S3, S4, S5를 추출한다. 최상위 특징에 해당하는

S5는AIFI(Attention-based Intra-scale Feature Interaction) Transformer 기반 인코더 층을 통해 연기를 잘 표현할 수 있는 특징 맵 F5로 변환되고, 다시{S3, S4, F5}는 MSFFPN(Multi-Scale Foreground-focus Fusion Pyramid Network)에 입력되어, 드론 영상에서 빈번히 발생하는 소형·희박 연기 영역에 집중하도록 설계된 전경 강화 과정을 거친 뒤P3, P4, P5 형태의 다중 스케일 피쳐로 융합된다. 이렇게 생성된 피쳐들은 평탄화 및 연결 과정을 통해 고정 길이의 시퀀스로 변환되고, IoU-aware query selection과deformable attention 기반 디코더로 구성된 검출 헤드를 거쳐 최종 산불 연기 검출 박스가 산출된다. 백본 단계에는 채널 중요도에 따라 연산 대상을 선택하는 부분 합성곱 모듈인 SCPConv와 효율적 다중 스케일 어텐션 모듈(EMA)이 결합되어 있어, 연산량을 줄이면서도 드론 영상 내 연기 특징에 대한 표현력을 향상시키도록 설계되었다.

2. Smoke Channel Partial Convolution (SCPConv)

연기 채널 부분 합성곱(SCPConv)은 기존 Partial Convolution(PConv)의 연산 효율성은 유지하면서도, 채널 선택 과정에 명시적인 기준을 도입해 드론 영상에서의 연기 표현력을 향상시키기 위해 설계된 모듈이다. 기존 PConv는 입력 특징 맵의 일부 채널에만 합성곱 연산을 수행함으로써 FLOPs를 줄이는 장점이 있으나, 어떤 채널을 선택할지에 대한 명확한 전략이 없어 정보 손실을 초래할 가능성이 존재한다. SCPConv는 채널 중요도 기반 선택 전략을 도입하고, 정보량이 큰 채널만 선별적으로 3×3 합성곱(RepConv)을 적용한다. 먼저 입력 특징 맵에 대해 글로벌 평균 풀링을 수행하여 채널별 통계 벡터를 계산한 뒤, 이를 두 개의 전결합(FC) 층과 sigmoid 함수를 통과시켜 채널 중요도 가중치를 산출한다. 이후 중요도 값이 상위에 해당하는 K개 채널의 인덱스를 선택 벡터로 정의하고, 이들 채널에 대해서만 RepConv를 수행한다. 선택되지 않은 나머지 채널은 1×1 합성곱, 배치 정규화(BN), ReLU로 구성된 저비용 경로를 통해 처리함으로써 전체 연산량을 크게 증가시키지 않으면서도 기본 표현력을 유지할 수 있다. 이를 통해 SCPConv는 드론 영상에서 연기와 같이 비정형적이고 저대비 특성을 갖는 중요한 채널에 보다 풍부한 비선형 변환을 제공하면서, 상대적으로 중요도가 낮은 채널은 경량 연산 경로로 유지하여 연산량 감소와 표현력 향상을 동시에 달성한다.

3. Efficient Multi-Scale Attention (EMA) and Backbone

효율적 다중 스케일 어텐션(EMA) 모듈은 채널과 공간 정보를 통합한 어텐션 구조로, 드론 영상에서 발생하는 연기의 불규칙한 형상과 넓은 공간 분포를 효과적으로 학습하기 위해 설계되었다. EMA는 우선 입력 특징맵 $X \in R(C \times H \times W)$ 를 R개의 채널 그룹으로 분할하고 각 그룹을 독립적으로 처리함으로써 연산량을 줄이는 동시에 다중 표현 학습을 가능하게 한다.

각 채널 그룹에 대해서는 두 개의 1×1 브랜치를 통해 각각 높이(H) 및 너비(W) 방향 평균 풀링을 수행함으로써 전역 방향 정보를 인코딩하고, 별도의 3×3 브랜치를 통해 국소 공간 구조를 추출한다. 이후 전역 및 국소 브랜치를 결합하여 채널 정보와 공간 정보를 동시에 반영한 통합 특징을 생성하며, 전역 특징과 3×3 공간 특징 사이에는 softmax와 행렬 곱 연산을 이용한 상호 어텐션(cross-spatial attention)을 수행해 전역 구조에서 세부 구조로, 세부 구조에서 전역 구조로 정보를 교환한다. 이를 통해 EMA는 드론 영상 내 연기의 확산 양상과 비국소적 패턴을 포착하는 데 유리한 표현을 형성하면서도 추가 파라미터 증가를 최소화한다.

백본 네트워크는 입력 영상을 대상으로 SCPConvBlock과 다운샘플링을 반복하는 4단계 층 구조로 구성되며, 이 중 마지막 세 단계에서 얻어지는 S3, S4, S5가 다중 스케일 특징으로 활용된다. SCPConvBlock은 ECPConv의 잔차 경로와 EMA를 결합함으로써, 연기와 직접적으로 연관된 채널 및 영역에 대한 응답을 선택적으로 강화하고 드론 영상 배경 잡음의 영향을 억제하는 역할을 수행한다.

4. Foreground-Focused Multi-Scale Pyramid (MSFFPN)

전경 집중 다중 스케일 피라미드 네트워크(MSFFPN: Multi-Scale Foreground-focus Fusion Pyramid Network)는 드론 영상에서 연기와 구름, 안개 등 배경 요소 간의 혼동을 줄이기 위해 다중 스케일에서 연기 전경 정보를 선택적으로 강조하는 것을 목표로 한다. MSFFPN의 핵심 구성 요소인 RCM(Rectangular Self-Calibration Module)은 직사각형 self-calibration 어텐션과 배치 정규화(BN), 다층 퍼셉트론(MLP)으로 구성된다.

RCM에서는 입력 특징에 대해 수평 및 수직 방향으로 각각 평균 풀링을 수행하여 두 방향의 전역 정보를 추출한 뒤, 이를 결합하여 전경 후보 영역을 모델링한다. 이후 $1 \times k$ 및 $k \times 1$ 형태의 비대칭 합성곱을 적용함으로써 드론

영상 내 가늘고 긴 연기 구조를 효율적으로 포착하며, 수평·수직 방향을 분리 처리함으로써 파라미터 수를 억제하면서도 비규칙적 형상에 유연하게 적응할 수 있도록 한다. 다중 스케일 전경 강화를 위해 백본에서 추출된 {S3, S4, S5}는 평균 풀링을 통해 동일한 해상도로 다운샘플링된 후 연결되어 피라미드 특징을 구성하고 RCM에 세 차례 통과시켜 다중 스케일 정보가 통합된 전경 강조 특징(P3-prime, P4-prime, P5-prime)을 생성한다. 식 (1)은 다중 스케일 전경 강조 특징 생성 과정을 나타낸다.

$$P_i' = RCM(Concat(AvgPool(S3), AvgPool(S4), AvgPool(S5))) \dots \dots \dots (식1)$$

Foreground Feature Fusion 단계에서는 식(2)를 이용해 각 스케일에서 원본 특징과 전경 강조 특징을 곱셈 및 덧셈 연산을 통해 상호 결합함으로써 전경(연기)에 대한 응답을 단계적으로 강화한다.

$$F_{out} = S3 \odot P3' \oplus S4 \odot P4' \oplus S5 \odot P5' \dots \dots \dots (식2)$$

최종적으로 P3-P5를 top-down 방식으로 융합하여 인코더 입력 벡터로 사용한다. 실험 결과, MSFFPN만 단독으로 적용한 경우에도 RT-DETR 대비 Precision이 3.1%p, Recall이 3.4%p, mAP50이 2.5%p 향상되는 것으로 나타나, 제안된 전경 집중 피라미드 구조가 드론 영상에서 연기와 배경을 분리하는 능력을 유의미하게 개선함을 확인하였다.

IV. Experiments and Evaluation

1. Experimental Environment and Dataset

실험 환경은 Intel® Xeon® E5-2680 v4 프로세서, 30 GB 시스템 메모리, 24GB VRAM을 갖는 NVIDIA GeForce RTX 3090 GPU로 구성된 클라우드 환경에서 수행하였다. 학습 프레임워크는 PyTorch 2.0을 사용하였으며, 배치 크기는 16, 초기 학습률은 0.001로 설정하고 cosine annealing 스케줄을 적용하여 150에포크 동안 학습을 진행하였다.

학습 및 평가에는 국내외 다양한 산림 환경에서 드론으로 촬영된 산불 연기 영상 데이터셋을 활용하였다. 본 데

이터셋은 총 12,500장의 고해상도 드론 영상 프레임(총 35,420개의 산불 연기 인스턴스)으로 구성되어 있으며, 다양한 고도(50~300m), 시간대(주간/야간/새벽), 기상 조건(맑음/흐림/안개)에서 촬영된 드론 영상 프레임을 포함한다. 전체 데이터셋은 모델의 객관적인 학습 및 평가를 위해 학습용 70%(8,750장), 검증용 15%(1,875장), 시험용 15%(1,875장)의 비율로 분할하였다. 드론 영상 특성에 맞는 데이터 증강 기법으로 랜덤 회전($\pm 30^\circ$), 스케일 변환(0.5~1.5 $\times$), 모션 블러, 밝기 조절 등을 적용하여 모델의 강건성을 확보하였다.

평가 지표로는 Precision, Recall, mAP50, mAP50-95(mAP95)를 사용하였다. 정밀도(Precision)의 TP는 연기로 올바르게 예측된 인스턴스 수, FP는 연기로 잘못 예측된 인스턴스의 수를 의미한다. 재현율(Recall)은 모델이 올바르게 예측한 모든 TP의 비율을 의미하며, FN은 모델이 연기로 예측하지 않았지만 실제 연기인 인스턴스 수를 의미한다. 식(3)은 Precision과 Recall의 수식을 나타낸다.

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN) \quad \dots\dots\dots (식3)$$

2. Performance Comparison with Baseline Models

표 1은 드론 영상 시험 세트에서 제안 모델과 주요 객체 검출기들을 비교한 결과이다.

Table 1. Performance Comparison of Object Detection Models

Model	Precision	Recall	mAP50	mAP95
Faster-RCNN	0.798	0.726	0.772	0.476
YOLOv11m	0.819	0.765	0.829	0.533
RT-DETR (baseline)	0.832	0.764	0.824	0.517
Proposed	0.883	0.800	0.862	0.539

그림 2는 실험결과를 시각화하여 나타낸 것으로 제안 모델은 RT-DETR 기준 모델 대비 Precision 0.7%p, Recall 2.4%p, mAP50 3.8%p, mAP95 2.2%p 향상을 달성하였다.

Faster R-CNN 대비로는 mAP50 기준 9.0%p의 큰 폭의 성능 향상을 보였으며, YOLOv11m과 비교해서도 mAP50 3.3%p, mAP95 0.6%p 향상되었다. 다만 mAP95(IoU 0.5-0.95 평균)는 다양한 IoU 임계값에서의 전반적 정밀 검출 성능을 나타내는 지표로, 본 실험에서는 객체 크기별(소형/중형/대형) 성능을 별도로 분리 측정하

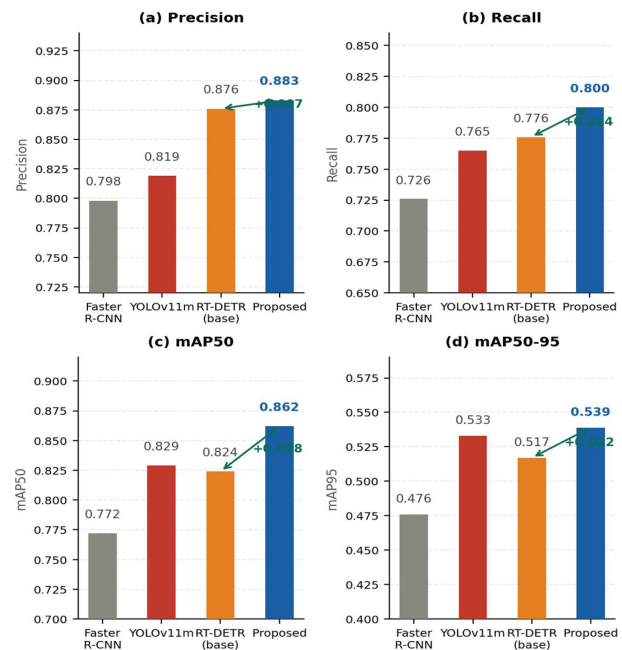


Fig. 2. Performance Comparison

지 않았으므로 "소형 연기 객체에서의 우수성"에 대한 직접적 근거로 사용하지 않으며, 이는 IV.5절 한계점에서 향후 분석 과제로 제시한다

3. Ablation Study and Cause Analysis

표 2는 제안 모델의 각 구성 요소의 기여도를 분석한 절제 실험(Ablation Study) 결과이다.

Table 2. Object Detection Results by Model

Model	Precision	Recall	mAP50	mAP95
RT-DETR-r18 (baseline)	0.832	0.764	0.824	0.517
+ PConv	0.847	0.769	0.832	0.520
+ SCPConv	0.851	0.776	0.843	0.521
+ SCPConv + EMA (SCPConvBlock)	0.851	0.791	0.852	0.535
+ SCPConvBlock + MSFFPN(Proposed)	0.883	0.800	0.862	0.539

표 2의 절제 실험 결과에 대한 심층 원인 분석은 다음과 같다. 첫째, 기본 RT-DETR에 PConv를 도입한 경우 연산 효율을 유지하면서도 Precision이 0.847로 상승하였으며, 여기에 채널 중요도 기반 선택 전략을 결합한 SCPConv를 적용한 결과 mAP50이 0.843으로 향상되었다. 둘째, SCPConv에 EMA(Efficient Multi-Scale Attention)를 추가하여 SCPConvBlock을 구성했을 때 Recall이 0.776에서 0.791로 크게 향상되었다. 셋째, 전경 집중 피라미드 네트워크(MSFFPN)를 최종 결합한 제안 모델은 Precision은 0.883, Recall은 0.800, mAP50은 0.862를

달성하여 각 구성 요소가 드론 연기 검출 성능 향상에 상보적으로 기여함을 객관적으로 입증하였다.

그림 3은 구성 요소별 성능 기여도를 도식화한 것으로 PConv를 SCPCConv로 교체하면 mAP50이 0.832에서 0.843으로 향상되어 채널 중요도 기반 선택 전략의 효과가 확인된다.

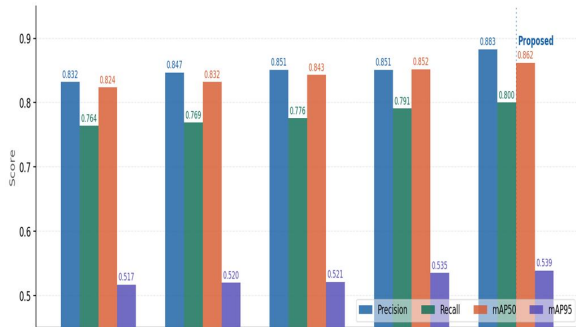


Fig. 3. Ablation Study Results by Component

EMA를 추가하여 SCPCConvBlock을 구성하면 Recall이 0.776에서 0.791로 크게 향상되는데, 이는 드론 영상 내 연기의 불규칙한 공간 분포를 EMA가 효과적으로 포착함을 시사한다. 최종적으로 MSFFPN을 결합한 완전한 제안 모델은 모든 지표에서 가장 우수한 성능을 달성하여 각 구성 요소의 상보적 기여를 확인하였다. 검증용 데이터셋에 대한 mAP50의 에포크별 변화를 분석한 결과 그림4와 같다.

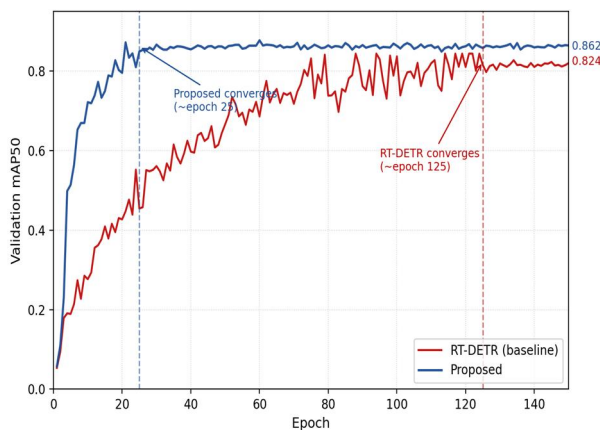


Fig. 4 Convergence comparison of validation mAP50 over training epochs

제안 모델은 학습 초반인 25 에포크 이전부터 안정적으로 수렴하는 것으로 나타났으며, RT-DETR은 약125 에포크가 지나서야 안정적인 수렴 양상을 보여 동일 조건에서 더 긴 학습 시간을 요구하였다. 이는SCPCConv의 채널 선택

전략과 EMA의 전역-국소 어텐션이 드론 영상의 연기 데이터로부터 유의미한 표현을 보다 빠르고 효율적으로 학습함을 의미한다. 학습 수렴 에포크 수 기준으로 RT-DETR 대비 약80% 감소하여 드론 탑재 또는 엣지 컴퓨팅 환경에서의 신속한 모델 업데이트에 유리함을 확인하였다.

본 연구의 실험 결과에서 드러난 한계로는 첫째, 특정 지형, 기상 조건, 카메라 배치 및 시야각등에 편향될 가능성이 있으며, 국내 산림·주거지·공장 등 다양한 환경에서 촬영된 드론 영상 데이터를 포함하여 재학습 및 검증을 수행함으로써 모델의 일반화 성능을 체계적으로 향상시킬 필요가 있다. 둘째, Transformer 기반 구조는 YOLO 계열 초경량 모델에 비해 여전히 연산량이 큰 편이기 때문에 드론 탑재 저전력 디바이스에서의 실시간 운용을 위해 네트워크 가지치기 및 지식 증류와 같은 추가적 경량화 기법의 도입이 필요하다. 셋째, 고정된 하이퍼파라미터와 난수 시드 하에서 단일 실행된 결과이므로 향후 반복 실행에 따른 통계적 변동성 검증이 요구된다.

V. Conclusions

본 논문에서는 드론 영상 기반 실시간 산불 감시를 위한 연기 특화 RT-DETR 개선 모델을 제안하였다. 채널 중요도 기반 선택적 합성곱(SCPCConv)과 효율적 다중 스케일 어텐션(EMA)을 결합한 SCPCConvBlock을 백본으로 구성하고, 전경 집중 다중 스케일 피라미드 네트워크(MSFFPN)를 통합함으로써 드론 영상에서 빈번히 발생하는 소형·희박 연기의 배경 혼동 문제를 효과적으로 해소하였다. 드론 촬영 산불 연기 데이터셋에서의 비교 실험 결과, 제안 모델은 RT-DETR 기준 모델 대비 Precision은 0.7%p, Recall은 2.4%p, mAP50은 3.8%p, mAP95는 2.2%p 향상을 달성하면서 학습 수렴 속도도 약 80% 단축하였다. 절제 실험을 통해 각 구성 요소의 기여도를 규명하였다. 향후 연구로는 멀티태스크 학습 기법 도입, Temporal Transformer 결합, 네트워크 가지치기 및 지식 증류를 통한 경량화를 계획하고 있다.

REFERENCES

- [1] Z. Ni, X. Chen, Y. Zhai, Y. Tang, and Y. Wang, "Context-Guided Spatial Feature Reconstruction for Efficient Semantic

- Segmentation," European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024. pp. 239-255, May 2024.
- [2] T. Shehzadi, K. A. Hashmi, M. Liwicki, D. Stricker, and M. Z. Afzal, "Object detection with transformers: A review," *Sensors*, vol. 25, no. 19, pp. 6025, October 2025.
- [3] J. Huang, J. Zhou, H. Yang, Y. Liu and H. Liu, "A small-target forest fire smoke detection model based on deformable transformer for end-to-end object detection," *Forests* 2023, vol. 14, no. 1, pp. 162, January 2023.
- [4] T. Liang, and G. Zeng, "FSH-DETR: An Efficient End-to-End Fire Smoke and Human Detection Based on a Deformable DETection TRansformer (DETR)," *Sensors* 2024, vol. 24, no. 13, pp. 4077.
- [5] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs Beat YOLOs on Real-time Object Detection," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 17-24 June 2024, pp. 16965-16974.
- [6] D. I. Lee, S. H. Son, J. G. Bae, S. Y. Park, S. H. Lee, J. M. Seo, Y. J. Kim, M. H. Choi, Y. W. Lee and J. S. Kim, "Enhancing Wildfire Damage Detection Performance with Swin Transformer Using Planetscope and Sentinel-2 Satellite Imagery: Effects of Multi-Source Data Fusion," *International Journal of Remote Sensing*, vol. 45, no. 6, pp. 991-1004, December 2024.
- [7] H. Sundaralingam, T. Suresh, T. Janith, and S. B. Ahmed, "Dilated Strip-Weighted Feature Pyramid: An Efficient Method for Object Detection," In *2025 IEEE International Symposium on Industrial Electronics (ISIE)*, pp. 1-6, June 2025.

Authors



Young-Bok Cho received the M.S., and Ph.D. degrees in Computer Science from Chungbuk National University, Korea, in 2003 and 2012, respectively. also Dr. Cho received more Ph.D degrees in Medical and

Law from Chungbuk National University and Chungnam National University, Korea, in 2019 and 2024, respectively. She has Professor of Information Security at Daejeon University, Daejeon, Korea, in 2018 to 2024, She is currently a Professor in the Computer Education at Gyeongbuk National University, Andong, Korea, in 2024. Her research interests include AI medical image processing, AI security and medical information protection, mobile security.