Journal of the KNST

2026; Vol.9, No.2; pp.391-397

ISSN: 2635-4926 https://doi.org/10.31818/JKNST.2026.6.9.2.391

Check for updates

Received: 2026/05/28
Revised: 2026/06/11
Accepted: 2026/06/21
Published: 2026/06/30

*Corresponding Author:
Jaehyun Bae
207 Mabuk-ro, Giheung-gu, YongIn-si, Gyeonggi-do, Republic of Korea
Tel: +82-31-326-9150
Fax: +82-31-326-9007
E-mail: jaehyun.bae@ligdna.com

적외선 영상에서의 표적 탐지 성능 향상을 위한 개선된 YOLOX 모델

An Improved YOLOX Model for Enhanced Target Detection in Infrared Images

배재현[1*], 김대현[2], 강병진[1], 백경훈[2]

[1]LIG디펜스&에어로스페이스 미사일시스템탐색기연구소 선임연구원
[2]LIG디펜스&에어로스페이스 미사일시스템탐색기연구소 수석연구원

Jaehyun Bae[1*], Daehyeon Kim[2], Byungjin Kang[1], Kyounghoon Baek[2]

[1]Research engineer, IIR Seeker R&D, LIG Defense&Aerospace
[2]Chief research engineer, IIR Seeker R&D, LIG Defense&Aerospace

Abstract

YOLOX는 일반 객체 검출에서 강력한 성능을 보여주지만, neck 구조가 주로 P3 - P5 수준의 저해상도 특성 맵에 의존하고 있어 작은 대상에 대한 검출 능력이 제한된다. 본 논문에서는 단순한 구조와 높은 추론 효율성을 갖춘 YOLOX를 베이스라인 모델로 선택했으며, neck 구조에 저수준 C2 특성 및 고수준 의미적 특징을 결합하는 방법을 제안한다. C2 특징이 제공하는 풍부한 공간 정보를 활용함으로써 거짓 양성을 효과적으로 감소시켰다. 제안한 수정된 구조는 기존 YOLOX 대비 $mAP_{50:95}$에서 최대 7.4%가 향상되었다.

Although YOLOX has demonstrated strong performance on generic object detection, its neck relies primarily on the low-resolution feature maps from levels P3–P5, which limits its ability to detect small targets. In this study, YOLOX was selected as the baseline model, and we propose a method which incorporates the low-level C2 features and high-level semantic features within the neck structure. By exploiting the richer spatial details of the C2 features, we effectively reduce false positives. The proposed modification improves detection performance by up to 7.4 % in $mAP_{50:95}$ compared to the baseline YOLOX.

Keywords

객체 탐지(Object Detection),
적외선 영상(Infrared Image),
특징 피라미드(Feature Pyramid),
딥러닝(Deep Learning),
합성곱 신경망(Convolutional Neural Network)

1. 서론

적외선 영상 기반의 표적 탐지, 추적 기술은 감시/정찰, 자율주행 등 다양한 분야에서 핵심 기술로 자리 잡고 있다. 이 중 감시/정찰 장비는 주·야간 연속 운용, 저조도, 안개, 기상 악화 등 다양한 운용 제약 조건에서 안정적인 표적 인지 능력 확보는 시스템 성능을 좌우하는 중요한 요소이다. 적외선 센서는 물체가 방사하는 열 에너지를 감지하기 때문에 외부 조명 조건에 대한 의존성이 낮고, 가시광 영상이 제한되는 환경에서도 신뢰성 있는 영상 정보를 제공할 수 있다. 이러한 특성으로 인해 적외선 영상은 군사, 국경, 해상, 무인 감시 플랫폼 등 다양한 국방 분야에서 핵심 센서로 활용되고 있다.

표적 탐지 및 추적은 감시정찰 시스템의 지능화 수준을 결정하는 핵심 기능이다. 탐지 단계에서는 관심 표적의 존재 여부와 위치를 정확하게 식별해야 하며, 추적 단계에서는 시간에 따른 표적의 운동 정보를 안정적으로 추정해야 한다. 그러나 적외선 영상은 가시광 영상과 비교할 때 텍스쳐 및 색상 정보가 제한적이며, 센서 해상도, 신호 대 잡음비(SNR), 배경 클러터(clutter) 등의 영향으로 인해 표적 식별이 어려운 특성을 가진다. 특히 소형 표적의 경우 배경 잡음과 쉽게 혼재되며 영상 내에서 잡음과 쉽게 혼재되며, 영상 내에서 표적이 차지하는 픽셀 수가 매우 적어 안정적인 탐지가 어렵다.

딥러닝 기반 객체 탐지 기술의 발전은 이러한 문제 해결에 중요한 전환점을 제공하고 있다. 합성곱 신경망(CNN)을 기반으로 한 데이터 중심 접근법은 복잡한 배경 조건 및 다양한 표적 특성에 대해 강인한 특징 표현을 학습할 수 있으며, 실시간 처리가 가능한 구조는 실제 감시/정찰 시스템 적용 가능성을 높인다. 따라서 적외선 영상 환경의 특성을 고려한 딥러닝 기반 탐지 알고리즘 설계는 학술적·실용적 측면에서 중요한 연구 주제로 부각되고 있다.

또한 딥러닝 기반 객체 탐지 모델은 대규모 데이터 학습과 고성능 연산 자원의 발전에 힘입어 비약적인 성능 향상을 이루어 왔다. 그 중 단일 단계(one-stage) 구조 객체 탐지기는 빠른 추론 속도와 비교적 단순한 구조로 인해 실시간 응용에 적합하며, YOLO(you only look once)[1-8] 계열 모델은 이러한 단일 단계 탐지기의 대표적인 예로 널리 활용되고 있다. 특히, YOLOX[9] 모델은 앵커 프리(anchor free) 설계를 도입하고, 효율적인 FPN(feature pyramid network)과 PAN(path aggregation network) 구조를 결합함으로써 정확도와 속도 측면에서 균형 잡힌 성능을 달성하였다.

YOLOX 모델 이외에 다른 SOTA(state of the art) 성능을 내는 많은 모델들이 존재한다. 그러나 이러한 모델을 방위산업 분야에 활용하기에는 몇 가지 제약이 존재한다. 방위산업 분야에서는 라이센스나 기술 제한, GPU 탑재 제한 등의 정책 및 운용적 제약이 존재한다. 또한, 최신 트랜스포머 기반 탐지기[10,11]는 큰 규모의 파라미터와 높은 연산량을 요구한다. 이는 저전력 임베디드 보드에서 실시간으로 구동하기 어렵다. 따라서, 본 논문은 방위산업 분야에서 임베디드 시스템 탑재 가능성을 핵심 목표로 삼아 YOLOX 모델을 선택하였다.

그러나 YOLOX를 포함한 다수의 객체 탐지 모델은 가시광 영상 데이터 셋을 중심으로 설계 및 학습되었으며, 소형 객체 또는 적외선 영상의 특성을 충분히 고려하지 않은 한계가 있다. 특히, YOLOX의 neck 구조는 P3-P5 수준의 다중 해상도 특징 맵을 활용하도록 설계되어 있어, 공간 해상도가 높은 P2 수준의 특징은 사용하지 않는다. 이로 인해 영상 내에서 매우 작은 영역을 차지하는 소형 표적에 대해 충분한 공간적 세부 정보를 제공하지 못하며, 이는 탐지 성능 저하에 영향을 미칠 수 있다.

본 논문에서는 적외선 영상에서의 소형 표적 탐지 성능 향상을 목표로 YOLOX의 neck 구조를 확장한 P2 기반 YOLOX 구조를 제안한다. 제안 기법은 기존 YOLOX의 PAN 구조에 C2와 P3 수준의 고해상도 특징을 통합함으로써, 소형 표적에 중요한 세부 특징을 효과적으로 활용한다. 적외선 소형 표적 데이터 셋인 anti-UAV 활용을 통해, 제안한 P2 기반 YOLOX 구조는 기존 YOLOX 대비 $mAP_{50:95}$(mean average precision)에서 최대 7.4%의 성능을 개선하였으며, 모든 AP 구간에서 성능 개선이 된 것을 확인하였다.

2. 관련 연구

2.1 적외선 소형 표적 탐지

적외선 영상에서의 소형 표적 탐지는 작은 픽셀 크기, 낮은 배경 대비, 제한적인 표적 특징 등의 특성으로 인해 어려운 문제로 알려져 있다. 전통적인 기법들은 영상 처리 기반의 수작업 특징 추출이나 필터 기반 기법을 이용해 표적을 추출하거나 배경 잡음을 억제하는 방식을 사용하였다. 하지만 복잡한 배경에서는 여전히 높은 오탐률과 낮은 검출률을 보였다.

딥러닝의 발전으로 인해 적외선 소형 표적 탐지에서도 CNN 기반의 모델이 도입되기 시작하였다. TBC-Net[12]은 적외선 소형 표적 탐지를 위해 경량 CNN 구조를 제안하였으며, 복잡한 배경에서 소형 표적 특징 학습 성능을 끌어올려 전통 기법 대비 성능이 크게 향상함을 보여주었다. iSmallNet[13]은 적외선 영상에서 표적과 배경 간의 불균형 문제를 해결하기 위해 ground-truth(GT) 레이블을 내부 영역과 경계 영역으로 분리하고, 작은 표적이 깊은 레이어에서 소실되지 않도록 다중 스케일 모듈을 적용하였다. 딥러닝과 CNN 기술의 발전으로 적외선 소형 표적 탐지 성능이 크게 향상되었으나, 여전히 작은 표적에 대한 탐지율과 복잡한 배경에서의 강인성에 대한 문제를 해결하지 못했다. 특히, 소형 표적은 매우 작아 깊은 레이어를 통해 전파되며 정보가 소실되기 쉽고, 특징도 국소적이기 때문에 일반적인 객체 탐지 네트워크 구조를 그대로 적용하기 어렵다.

2.2 YOLO 기반 적외선 영상 객체 탐지

YOLO 계열 모델은 단일 단계 객체 탐지기의 대표

모델로 빠른 속도와 효율성으로 실시간 탐지 응용에 널리 사용되고 있다. 특히 적외선 영상처럼 저해상도 및 낮은 대비 환경에서 YOLO 기반의 모델을 활용하여 소형 객체 탐지를 수행하는 연구가 활발히 진행되고 있다.

딥러닝 모델의 다운샘플링 과정에서 소형 표적 정보가 손실되는 문제를 완화하기 위해 YOLO-FR[14] 모델이 제안되었다. 해당 모델은 YOLOv5를 기반으로 특징 재병합 샘플 방법을 도입하여 적외선 소형 표적 검출 성능을 기존 모델 대비 크게 개선하였다.

IRST-YOLO[15]는 YOLOv8n을 기반으로 세분화된 다운샘플링 모듈과 어텐션 기법을 적용하여 적외선 소형 표적에 대한 검출 정밀도를 향상시켰다. YOLO-UIR[16]은 UAV 기반 적외선 객체 탐지에 특화된 경량화된 YOLO 모델을 개발하여 작은 객체의 표현력을 향상시켰다.

이처럼 YOLO 기반 연구들은 백본 및 넥 구조의 개선, 어텐션 모듈 통합과 경량화 등을 통해 적외선 영상 도메인에 맞는 모델을 개발하기 위해 발전해왔다. 그러나 이러한 연구에서도 일반적으로 P3 이상 레벨 중심의 스케일만 활용하거나 연산 비용을 증가시키는 구조가 많아, 네트워크 효율성과 소형 표적 특징 간의 균형을 고려하는 연구는 여전히 필요하다.

3. 실험 과정 및 성능 평가

3.1 데이터 셋 및 실험 방법

적외선 영상 기반의 소형 데이터 셋을 직접 획득하기는 굉장히 어렵기 때문에 본 논문에서는 공개 데이터 셋 중 소형 드론을 촬영한 AntiUAV[17] 데이터를 사용하였다. Fig. 1은 AntiUAV 데이터 셋의 일부이다. 적외선 영상 기반의 단일 객체가 포함된 영상이다. 해당 데이터 셋은 챌린지에서 공개되었으며, 챌린지의 특성상 테스트 셋은 레이블링 데이터가 공개되지 않는다. 실험에 해당 영상을 활용하기 위해 학습, 검증 데이터를 혼합하여 학습, 검증, 테스트를 8:1:1의 비율을 갖도록 재분배하였다. 이때, 영상 시퀀스 별로 분배하였기 때문에 학습에 사용한 시퀀스가 테스트에 등장하지 않는다. 또한, 영상 내 드론이 존재하지 않는 영상은 모두 제외하였으며, 총 227,435장이 사용되었다. 영상의 해상도는 640×512이며, 영상 전처리에 사용된 기법과 카메라의 파장 대역 등의 정보들은 공개되지 않았다.

Fig. 1. A sample image of the AntiUAV dataset

본 논문에서 사용한 모델은 YOLOX 중 small(S)와 nano 두 가지의 크기를 가진 모델을 활용하였다. 학습 환경은 배치 크기 16으로 설정하였으며, mix up 기법은 사용하지 않았다. 입력 해상도는 640×640이며, 학습 에폭은 300, warm up 에폭은 5, 초기 학습률은 0.003으로 변경하였다. 학습 및 추론은 RTX 4090에서 수행되었다.

Fig. 2는 본 논문의 프레임워크를 도식화하였다. 기존 YOLOX 모델은 백본으로 영상이 입력된 후 여러 특징맵 단계인 C3에서 C5 단계의 특징 맵을 neck에 입력하여 다중 스케일 특징을 생성하고, 헤드에서 이를 이용해 객체 탐지를 수행한다. 이때, 소형 객체는 고해상도 특징 맵에서 사라지는 경향이 있다. Table 1은 입력 영상의 해상도가 640×640인 영상에서 타겟의 크기가 16×16인 경우 특징 맵에서 크기가 어떻게 변화하는지 나타내었다. 기존 YOLOX 모델은 C3에서 C5만 사용하였기 때문에 2×2로 크기가 줄어서 타겟의 최소한의 모양 표현도 어렵다. 본 논문이 제안하는 기법은 C2 특징 맵을 사용하는 것이다. C2 특징 맵은 더 높은 해상도를 갖고 있기 때문에 작은 객체가 특징 맵에서 사라지지 않고, 소형 표적의 위치 정밀도가 증가한다는 장점이 있다. 본 논문에서는 백본에서 추출한 C2 특징 맵을 P2로 변환하고, P2를 PAN에서만 사용하도록 구현하였다. 구체적으로 P3 특징을 업샘플링한 후, C2와 결합하여 보폭 4 해상도

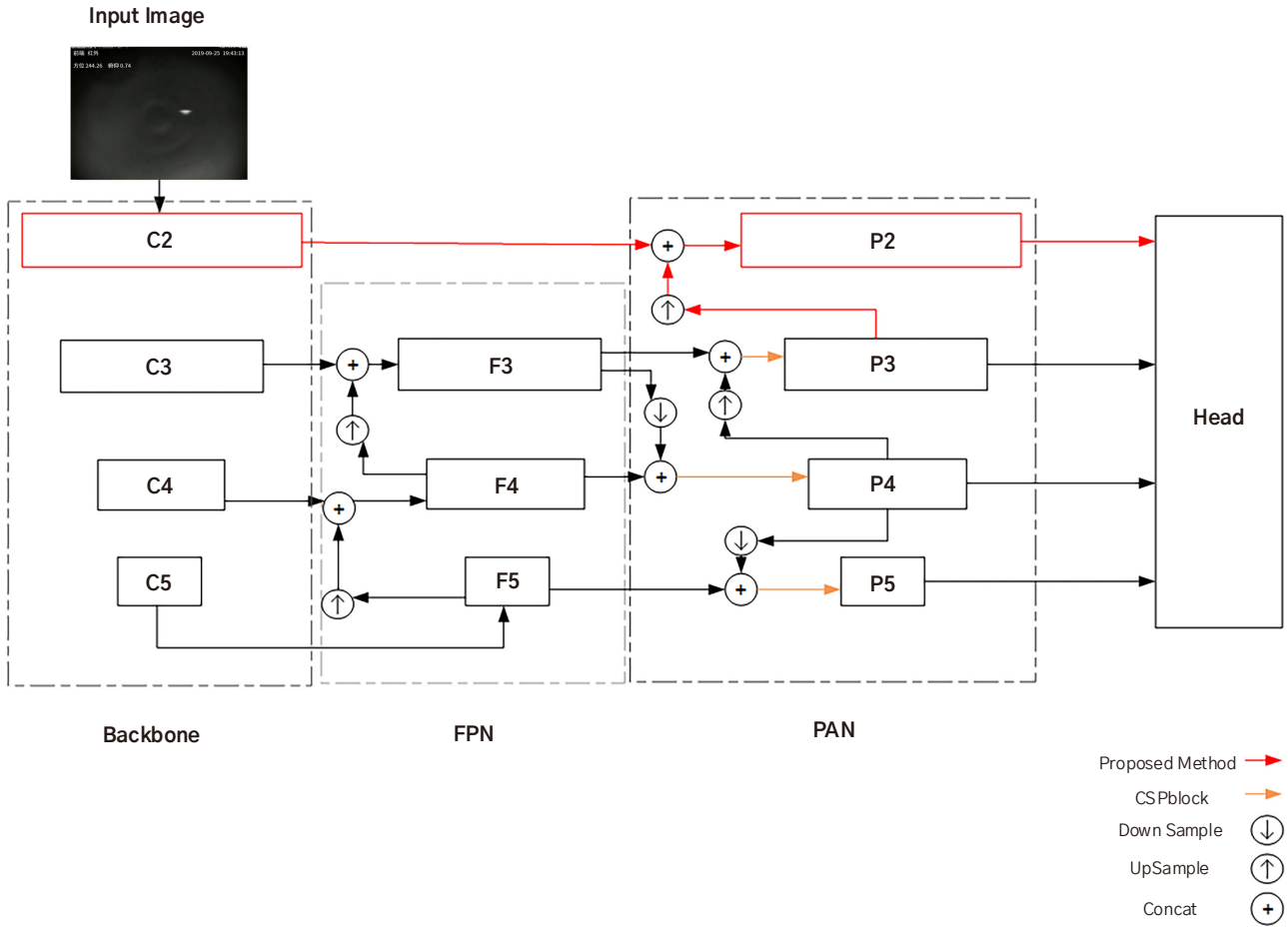

Fig. 2. The Framework of the proposed method. Model is consisted of backbone, FPN, PAN, and head, which is same as the original YOLOX model. The original model utilized the feature maps from C3 to C5. Our proposed method (red) incorporates the C2 features in both neck and head to enhance the multi-scale feature

의 P2 특징을 생성하였다. 생성된 P2 특징은 head에 직접 입력되며, PAN에서 bottom-up 경로에는 추가적으로 사용하지 않았다. 이는 추가적인 연산량 및 복잡도 증가를 야기하여 추후 임베디드 실시간성을 저해하기 때문이다.

Table 2는 기존 YOLOX 모델과 P2-YOLOX 모델의 파라미터 수와 연산량을 각각 표시하였다. 파라미터 수 대비 연산량이 크게 증가하는데, 이는 고해상도인 C2와 P2 특징 맵을 사용하며 3×3의 컨볼루션 연산에 의해 계산 복잡도가 증가하였기 때문이다. 추론 시간 또한 기존 모델 대비 약 2배가 되었으나 방위산업 분야에서는 연산량 증가의 문제도 중요하지만, 표적을 망실하는 것이 더 심각한 문제이며, 임베디드화 과정에서 TensorRT 등의 최적화를 통해 추론 속도의 경우 개선 가능성이 여전히 존재한다.

Table 1. The target sizes computed for each feature map stage at the corresponding resolutions

Feature map	Resolution	Stride	Target size
-	640 x 640	-	16 x 16
C2 / P2	160 x 160	4	4 x 4
C3 / P3	80 x 80	8	2 x 2
C4 / P4	40 x 40	16	1 x 1
C5 / P5	20 x 20	32	0.5 x 0.5

Table 2. The number of parameters, FLOPs and inference time of baseline YOLOX and the proposed P2-YOLOX

	YOLOX-S	P2-S	YOLOX-Nano	P2-Nano
Params (M)	8.94	9.5	2.24	2.4
FLOPS (G)	26.76	58.8	6.93	15.01
Inference Time (ms)	1.28	2.59	0.59	1.07

Table 3. Comparison of the AP, recall, and FP/TP ratio for different object sizes between the baseline YOLOX and the proposed P2-YOLOX models. "P2-X" denotes the architecture and model size of the proposed YOLOX model. The better value for each mode and metric is highlighted in bold. s denotes small size and m denotes medium size and the mean average precision (*mAP*) of 50 and 50:95 is written below

	YOLOX-S	P2-S	YOLOX-Nano	P2-Nano
$AP_{s@50}$ | recall | FP/TP	0.383 | **0.864** | 4.42	**0.405** | **0.864** | **3.31**	0.385 | **0.882** | 8.46	**0.392** | 0.862 | **4.31**
$AP_{m@50}$ | recall | FP/TP	0.368 | **0.991** | 3.07	**0.401** | 0.990 | **2.26**	0.514 | **0.988** | 5.53	**0.526** | 0.984 | **2.53**
$AP_{s@60}$ | recall | FP/TP	0.298 | 0.811 | 4.72	**0.373** | **0.828** | **3.46**	0.353 | **0.832** | 8.78	**0.359** | 0.815 | **4.54**
$AP_{m@60}$ | recall | FP/TP	0.351 | **0.967** | 3.16	**0.395** | 0.966 | **2.33**	0.508 | **0.968** | 5.63	**0.521** | 0.967 | **2.58**
$AP_{s@70}$ | recall | FP/TP	**0.233** | 0.638 | 6.14	0.195 | **0.688** | **4.28**	0.28 | 0.654 | 11.31	**0.283** | **0.660** | **5.71**
$AP_{m@70}$ | recall | FP/TP	**0.314** | **0.768** | 4.19	0.241 | 0.765 | **3.16**	0.454 | 0.794 | 7.05	**0.474** | **0.81** | **3.25**
$AP_{s@80}$ | recall | FP/TP	**0.082** | 0.305 | 13.31	0.038 | **0.312** | **10.07**	**0.103** | 0.236 | 32.88	0.031 | **0.256** | **15.72**
$AP_{m@80}$ | recall | FP/TP	**0.109** | **0.28** | 12.82	0.034 | **0.28** | **9.98**	**0.21** | 0.315 | 19.04	0.078 | **0.322** | **9.49**
$AP_{s@90}$ | recall | FP/TP	0.004 | 0.007 | 602	**0.005** | **0.01** | **317**	0.001 | 0.002 | 3198	**0.002** | **0.004** | **860**
$AP_{m@90}$ | recall | FP/TP	**0.002** | 0.004 | 959	0.001 | **0.06** | **502**	**0.009** | **0.008** | 799	0.006 | 0.005 | **644**
mAP_{50}	0.706	**0.735**	**0.741**	0.739
$mAP_{50:95}$	0.377	**0.405**	0.417	**0.428**

Table 3는 객체의 크기(s, m) 별 AP, recall, FP/TP 비율을 기존 YOLOX 모델과 제안하는 YOLOX 구조의 mAP를 구간별로 나열하였다. 이때, 객체의 크기를 s와 m으로 나누는 기준은 32*32보다 작은 경우 s(small), 96*96보다 작은 경우 m(medium)으로 구분하는 COCO 데이터 셋의 기준을 따라가였다. 총 24,546장의 테스트 셋 중에 large 객체의 경우 1,739장 밖에 포함되지 않아 유의미한 결과를 나타내지 못하여 해당 실험에서 제외하였다. AP의 경우, P2를 적용하는 것이 S 모델에서는 5구간, Nano에서는 7구간의 객체 크기에서 유리한 것을 확인할 수 있다. Recall의 경우, S 모델에서는 약간 도움이 되었으나 Nano 모델에서는 큰 개선점이 없었다. 하지만 FP/TP 비율의 경우 P2를 적용했을 때 모든 구간에서 안정적으로 감소한다. 이러한 결과를 통해 P2 특징 맵이 단순히 검출 개수를 늘린다기보다 불필요한 검출을 줄이며 예측 품질을 개선했음을 의미한다. mAP_{50}의 경우, Nano 모델에는 0.2%p가량 하락하였으나 S 모델에서는 2.9%p 개선되었으며, 전 구간에 대한 성능을 나타내는 지표인 $mAP_{50:95}$에서는 모두 성능이 개선되었다.

Fig. 3는 기존 모델과 제안 기법을 적용한 모델의 탐지 결과를 나타내었다. S 모델과 Nano 모델 모두에서 기존 대비 제안 기법을 적용한 P2 모델에서 탐지 신뢰도가 높아졌으며, 특히 탐지하지 못했던 소형 표적도 제안 기법을 적용한 P2-Nano에서 성공적으로 탐지하는 것을 확인하였다. 고해상도 특징 맵인 P2를 활용함으로써 소형 표적에 대한 공간 정보를 보다 효과적으로 유지할 수 있기 때문으로 판단된다. 그러나 이러한 결과만으로는 P2 특징 맵이 실제로 소형 표적에 대한 어텐션을 강화하는지 직관적으로 확인하기에는 한계가 있다.

따라서 제안한 P2 특징 맵이 소형 표적 영역에 어떠한 영향을 미치는지 정성적으로 분석하기 위해 Grad-CAM[18] 시각화를 수행하였다. Grad-CAM 분석은 제한된 특징 표현 능력을 가지는 경량 모델 환경에서 P2 특징 맵의 효과를 보다 명확하게 확인할 수 있는 Nano 모델을 대상으로 수행하였다. 기존 YOLOX 모델에서 가장 높은 해상도를 가지는 P3 특징 맵과 제안 모델의 P3 특징 맵에 대한 활성화 영역을 비교하였다. Fig. 4는 Nano 모델에서 기존 모델과 제안 기법을 적용한 모델의 Grad-CAM 결과를 보여준다. Nano 모델의 P3 특징에서는 저대비, 소형 표적 영역이 활성화되지 못하는 반면, P2-Nano의 P2 특징에서는 표적 주변 영역이 활성화되는 것을 확인할 수 있다. 이는 P2 특징 맵이 표적의 공간적 특징을 보다 효과적으로 유지하고 있음을 의미한다.

실험을 통해 적외선 영상 기반의 객체 탐지 문제에서 기존 YOLOX 모델 대비 P2 특징 맵을 활용하는 것이 recall 자체를 크게 개선하지 않지만 오탐지를 획기적으로 개선하여 탐지 성능 개선에 기여하는 것을 확인하였다. 모델의 크기와 관련 없이 $mAP_{50:95}$ 측면

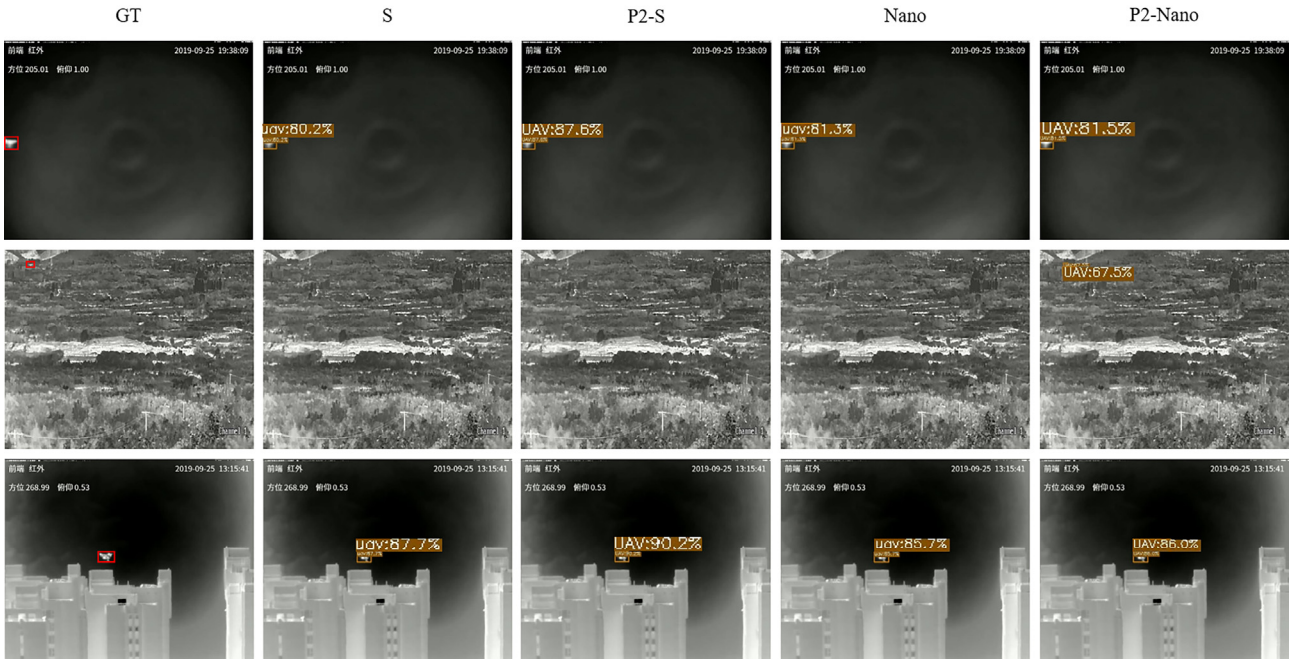

Fig. 3. The comparison of detection result of the variants of YOLOX models. Red box shows the ground-truth(GT) and the orange box shows the detection results with the confidence score

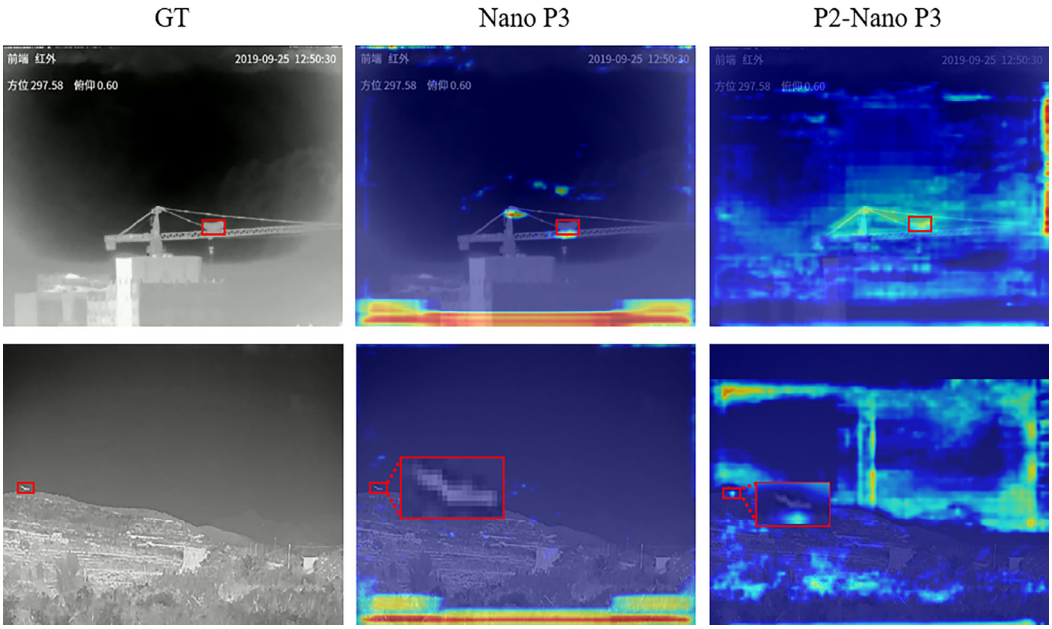

Fig. 4. The comparison of the result of Grad-CAM visualization. Red box shows the activation near the object

에서 PAN 과정에서 P2 특징 맵을 활용했을 때 기존 모델 대비 성능이 개선됨을 실험을 통해 검증하였다.

4. 결론

본 논문에서는 방위산업 분야에서 임베디드 탑재 가능성을 고려하여 적외선 영상 기반의 탐지 알고리즘 중 YOLOX 모델을 선정하여 실시간으로 구동할 수 있도록 탐지 성능을 개선하는 것을 목표로 하였다. 기존 YOLOX 모델은 백본에서 C3부터 C5에 해당하는 특징 맵을 사용하여 FPN과 PAN에서 특징 융합을 수행하였다. 하지만 C3에 입력되는 해상도에서

소형 객체의 경우 크기가 매우 작아져 공간 정보를 사용하지 못하는 경우가 있다. 본 논문에서는 YOLOX 모델의 탐지 성능을 개선하기 위한 연구를 수행하였으며 그 결과는 다음과 같다.

(1) C2 특징을 사용하여 공간정보를 사용하여 PAN에서 이를 P3와의 융합을 통해 P2 특징을 생성하였다.

(2) P2 특징 사용을 통해 재현을 크게 개선하기보다 오탐지를 줄여 탐지 성능을 개선하였다.

향후 연구에서는 기존 모델과 제안한 P2 기반 모델을 임베디드 보드에 포팅하여 실 구동 시의 레이턴시와 최적화 기법 적용을 통한 모델의 성능을 분석할 예정이다.

참고문헌

[1] Joseph Redmon, Santosh Divvala, Ross Girshick, & Ali Farhadi, 'You Only Look Once: Unified, Real-Time Object Detection,' in proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779-788.
[2] Joseph Redmon & Ali Farhadi, 'YOLO9000: Better, Faster, Stronger,' in proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7263-7271.
[3] Joseph Redmon & Ali Farhadi, 'YOLOv3: An Incremental Improvement,' arXiv preprint arXiv:1804.02767, 2018.
[4] Alexey Bochkovskiy, Chien-Yao Wang, & Hong-Yuan Mark Liao, 'YOLOv4: Optimal Speed and Accuracy of Object Detection,' arXiv preprint arXiv:2004.10934, 2020.
[5] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, Yonghye Kwon, Kalen Michael, Jiacong Fang, Lorna Imyhxy, Colin Wong, Yifu Zeng, et al. (2022). 'ultralytics/yolov5: v6. 2-YOLOv5 Classification Models, Apple M1, Reproducibility, ClearML and Deci.ai Integrations,' Zenodo.
[6] Chien-Yao Wang, Alexey Bochkovskiy, & Hong-Yuan Mark Liao, 'YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors,' in proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7464-7475.
[7] Rejin Varghese & M. Sambath, 'YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness,' in proceedings of 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, April 2024, pp. 1-6.
[8] Chien-Yao Wang, I-Hau Yeh, & Hong-Yuan Mark Liao, YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information, In Computer Vision - ECCV 2024, Springer Nature Switzerland, September 2024, pp. 1-21.
[9] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, & Jian Sun, 'YOLOX: Exceeding YOLO Series in 2021,' arXiv preprint arXiv:2107.08430, 2021.
[10] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M. Ni, & Heung-Yeung Shum, 'DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection,' arXiv preprint arXiv:2203.03605, 2022.
[11] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, & Jie Chen, 'DETRs Beat YOLOs on Real-Time Object Detection,' in proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 16965-16974.
[12] Mingxin Zhao, Li Cheng, Xu Yang, Peng Feng, Liyuan Liu, & Nanjian Wu, 'TBC-Net: A Real-Time Detector for Infrared Small Target Detection Using Semantic Constraint,' arXiv preprint arXiv:2001.05852, 2019.
[13] Zhiheng Hu, Yongzhen Wang, Peng Li, Jie Qin, Haoran Xie, & Mingqiang Wei, 'ISmallNet: Densely Nested Network with Label Decoupling for Infrared Small Target Detection,' in proceedings of ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, June 2023.
[14] Xingang Mou, Shuai Lei, & Xiao Zhou, 'YOLO-FR: A YOLOv5 Infrared Small Target Detection Algorithm Based on Feature Reassembly Sampling Method,' Sensors, VOL. 23, NO. 5, 2023, Article 2710.
[15] Huicong Wang, Kaijun Ma, Juan Yue, Yuhan Li, Jiaxin Huang, Jie Liu, Linhan Li, Xiaoyu Wang, Nengbin Cai, & Sili Gao, 'Small-Target Detection Based on Improved YOLOv8 for Infrared Imagery,' Electronics, VOL. 14, NO. 5, 2025, Article 947.
[16] Chao Wang, Rongdi Wang, Ziwei Wu, Zetao Bian, & Tao Huang, 'YOLO-UIR: A Lightweight and Accurate Infrared Object Detection Network Using UAV Platforms,' Drones, VOL. 9, NO. 7, 2025, 479.
[17] Nan Jiang, Kuiran Wang, Xiaoke Peng, Xuehui Yu, Qiang Wang, Junliang Xing, Guorong Li, Jian Zhao, Guodong Guo, & Zhenjun Han, 'Anti-UAV: A Large-Scale Benchmark for Vision-Based UAV Tracking,' IEEE Transactions on Multimedia, VOL. 25, November 2021, pp. 486-500.
[18] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, & Dhruv Batra, 'Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,' in proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618-626.