

# YOLOv8-P2: 고해상도 특징 기반 객체 탐지 모델

김윤희, 서민경, 현장훈\*, 김영동\*  
국립한밭대학교\*

20221047@edu.hanbat.ac.kr, [20221052@edu.hanbat.ac.kr](mailto:20221052@edu.hanbat.ac.kr),  
[jhhyeon@hanbat.ac.kr](mailto:jhhyeon@hanbat.ac.kr)\*, ydkim1293@hanbat.ac.kr\*

## YOLOv8-P2: High-Resolution Feature-Based Object Detection Model

Yunhee Kim, Minkyoung Seo, Janghun Hyeon\*, Youngdong Kim\*  
Hanbat National University\*

### 요약

본 연구는 철도 환경에서 원거리의 작은 객체 탐지 성능을 향상시키기 위한 객체 탐지 모델 구조 개선 방법을 제안한다. 기차 및 트램은 긴 제동거리로 인해 먼 거리의 객체를 조기에 인식해야 하며, 원거리 객체는 영상에서 매우 작은 크기로 표현되어 기존 탐지 모델의 성능 저하가 발생한다. 이를 해결하기 위해 YOLO v8 모델을 기반으로 기존 P3 부터 수행되던 객체 탐지를 P2 계층까지 확장하는 Detection Head 를 추가하여 고해상도 특징을 활용한 멀티스케일 탐지 구조를 제안한다. 실험 결과, 제안 모델은 원거리 객체의 형태 정보를 효과적으로 보존하여 기존 구조 대비 탐지 성능 향상을 보였으며, 향후 데이터 확장을 통해 추가적인 성능 개선이 가능할 것으로 기대된다.

### I. 서론

기차, 트램과 같은 운송수단은 주행 속도가 높고 무게가 무거워 제동거리가 매우 길다. 따라서 먼 거리의 객체를 조기에 탐지하는 기술이 필요하다. 예를 들어 시속 150 km 로 주행하는 승용차의 경우 제동거리가 100~150m 이지만 기차의 경우 약 1000m 를 달한다. 또한 기차는 레일을 따라 움직이기에 경로에 제한이 있다. 이러한 특성으로 인해 갑자기 객체가 뛰어나오는 돌발 상황 발생 시 즉각적인 대응이 어려워 수백미터 이상 떨어진 객체를 효과적으로 인지하는 탐지 기술이 중요하다.

그러나 실제 원거리 객체는 영상 내에 매우 작은 크기로 표현되어 일반적인 객체 탐지 모델에서는 한계가 존재한다. 특히 철도 환경이라는 특수성으로 인해 데이터 규모가 제한적이며 기존의 물체 탐지 기술로는 충분한 성능을 기대할 수 없다. 본 연구는 이러한 문제를 해결하기 위해 기존 객체 탐지 모델(YOLO v8)의 구조를 수정하여 원거리 객체 인식의 성능을 높이는 방법을 제안한다.

### II. 본론

본 연구에서는 기존 YOLO v8[2] 모델을 Base 로 잡고 구조를 수정한다. 모델 수정에 앞서 이미지 데이터를 직접 라벨링하였으며 해당 결과를 GT(Ground Truth)로 설정하고 학습을 진행하였다. 기존의 YOLO v8 모델은 이미지가 1/8 배 된 시점(P3) 부터 탐지를 수행한다. 하지만 이는 이미 작은 객체가 다운샘플링 과정에서 형태 정보가 소실된다. 작은 객체가 더 작아져 정보가 소실되는 것이다. 이러한 문제를 해결하기 위해 기존 P3 부터 시작하던 객체 탐지를 이미지가 1/4 배 되는 지점인 P2 부터 진행하도록 Detection Head 를 추가한다. 이를 추가하여 기존 세 단계의 멀티스케일

탐지 구조(P3, P4, P5)를 네 단계의 멀티스케일 탐지 구조(P2, P3, P4, P5)로 구성한다.

#### 2.1 데이터 라벨링

모든 데이터는 한국철도기술연구원에서 제공한 이미지 자료를 기반으로 한다. 이미지와 3D Box 값이 채워져 있는 KITTI 포맷 라벨, 카메라 캘리브레이션 좌표값이 존재한다. 이를 기반으로 이미지에 3D Box 의 값을 카메라 캘리브레이션 좌표값을 기반으로 투영시켜 KITTI 포맷의 2D Bounding Box 라벨을 생성하였다.



그림 1. 라벨링 결과

이러한 이미지는 총 240 장으로 학습용 데이터셋을 구축하였다. 철도 선로 전방에서 촬영된 실제 영상을 각 초 단위로 나눈 이미지 데이터셋은 선로 상의 인물, 차량, 장애물 등 다양한 객체가 포함되어 있다.

#### 2.1 제안 모델

기존 YOLO v8 모델의 경우 이미지가 1/8 배 축소된 시점(P3)부터 탐지를 수행한다. 이는 원거리의 작은 사이즈로 나타나는 객체가 형태 정보를 소실하게 되어 탐지에 어려움을 겪는다. 본 연구에서는 앞선 문제를 해결하기 위해 이미지가 1/4 배 되는 시점(P2)부터 객체 탐지를 수행하기 위해 Detection Head 를 P2 에 추가하여 모델의 구조를 확장한다.

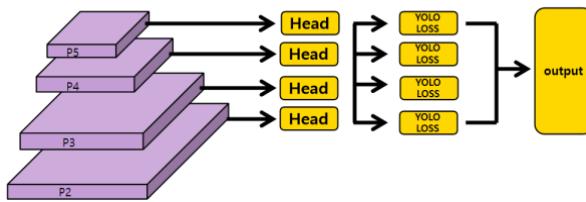


그림 2. 제안 모델 구조

Base Model 이 되는 YOLO v8 의 P2 는 객체 탐지는 수행하지 않고 feature map 을 추출하여 P3 가 더 좋은 정보를 전달받을 수 있게 하는 역할이었다. 본 연구에서는 작은 객체를 탐지하기 위해 이미지의 사이즈가 너무 작아지기 전에 객체 탐지를 수행해야한다고 판단했다.

Detection Head 는 기존 세 단계(P3, P4, P5) 였던 YOLO v8 에 P2 Detection Head 를 추가하여 네 단계의 멀티스케일 탐지 구조(P2, P3, P4, P5)를 구성하였다. P2 Head 는 고해상도의 이미지 입력을 기반으로 추출된 feature map 을 기반으로 탐지한다. 최종 예측을 위해 각 네 단계의 해상도에서 출력된 결과를 통합하여 최종 예측을 진행한다. 이런 구조적 확장은 고해상도의 특징을 활용하여 Detection Head 에 입력으로 사용하기 때문에 작은 물체에 더 민감하게 반응하여 원거리 물체 탐지 성능을 안정적으로 유지할 수 있다.

### III. 실험 결과 및 분석

본 연구의 모든 실험은 NVIDIA RTX A6000 GPU 환경에서 진행하였고, 한국철도기술연구원의 철도 주행 영상을 기반으로 만든 이미지를 학습에 총 240 장, 검증에 43 장 사용하였다.

실험 모델은 총 3 개이며, 사전학습된 YOLO v8 모델, 증강 데이터를 통해 파인튜닝 모델, 제안 모델이다. YOLO v8 모델이 Base Model 이며, 사전학습 모델에 데이터를 증강하여 파인튜닝한 모델이 두번째 증강 모델이다. 이때 사용한 데이터 증강 기법으로는 이미지의 밝기/대비 조절, 이미지의 흐림정도를 조절하는 가우시안 블러, 디지털 노이즈를 추가하는 가우시안 노이즈, 색도와 채도, 명도를 변경하는 HSV 색공간변형을 사용하였다. 최종 제안 모델로는 사전학습 모델에 P2 Detection Head 를 추가하여 증강데이터로 파인튜닝하였다.

	Accuracy	Recall	Precision	mAP50	mAP50-95
Base Model	0.233	0.163	0.412	0.248	0.034
증강 Model	0.954	0.953	0.953	0.97	0.631
제안 Model	0.933	0.953	0.953	0.97	0.607

표 1. 성능 표

Base Model 은 Accuracy 0.233, Recall 0.163, Precision 0.412, mAP50- 95 0.034 로 전반적으로 매우 낮은 성능을 보인다. 이는 앞서 언급했던 실험 환경의 특수성과 기본 구조인 P3 ~P5 계층 만으로 객체 정보를 충분히 보존하지 못했기 때문이다. 반면 증강 Model 은 다양한 증강 기법을 사용하고 데이터를 파인튜닝하여 Accuracy 0.954, Recall 0.953, Precision 0.953,

mAP50- 95 0.631 로 Accuracy 기준 약 0.7 의 성능을 향상시켰다. 제안 Model 은 Accuracy 0.933, Recall 0.953, Precision 0.953, mAP50- 95 0.607 로 증강 Model 과 수치상으로 유사하지만, 실제 검출 결과에서 약 0.02 정도 미세하게 성능이 하락하는 모습을 보였다.



그림 3. Base Model 예측 결과

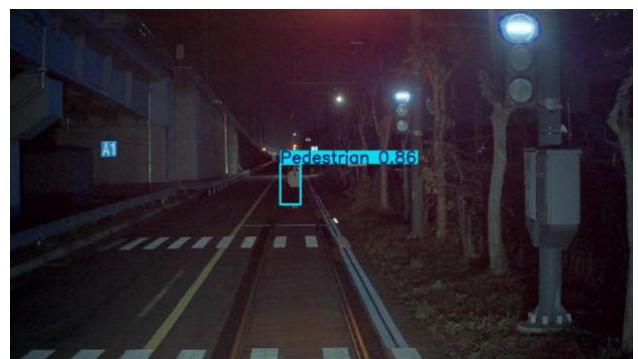


그림 4. 제안 모델 예측 결과

### IV. 결론

본 연구에서는 원거리에 있는 작은 물체 탐지 성능을 향상시키기 위해 YOLO v8 모델 구조에 P2 Detection Head 를 추가하였다. 제안 Model 은 기존 YOLO v8 대비 객체의 형태 정보를 잘 보존하여 높은 성능 향상을 보여주었다. 실험결과, 제안 Model 이 증강 Model 에 비해 다소 낮은 값을 기록하였다. 이는 학습 및 검증 데이터의 규모가 제한적이라 P2 계층이 충분한 학습을 하지 못한 것으로 판단된다. 따라서 향후 연구에서는 충분한 데이터셋의 양과 다양성을 확보하여, P2 계층을 충분히 학습시킨면 제안 구조의 성능을 더욱 향상시킬 수 있을 것으로 기대한다.

### 참 고 문 헌

- [1] T. Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017

- [2] Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016.