

# VisDrone-DET 학습데이터를 활용한 YOLO 버전별 분류 정확도 변화 비교 연구

## Comparison of changes in classification accuracy by YOLO version using VisDrone-DET training data

최혁주<sup>1)</sup> · 윤준희<sup>2)</sup> · 유수홍<sup>3)</sup>

Choi, Hyeokjoo · Youn, Junhee · Yoo, Suhong

### Abstract

Object detection is actively used in various fields of society, such as autonomous driving, safety management, urban management, smart factories, and environmental protection. Meanwhile, as multiple versions of YOLO are announced, more research is needed on selecting models and hyperparameters suitable for training data. Therefore, in this study, the performance of various versions of YOLO (v5, v6, v8, v10, and v11) was verified through experiments using VisDrone-DET training data. In addition, the change in accuracy was analyzed by changing the weight size, input image resolution, optimization function, and learning rate among the hyperparameters. As a result of the experiment, the YOLO v8 XLarge model showed the highest accuracy of 0.669 based on mAP@50 when the Adam optimization function, learning rate of 0.001, and input image size of 1920 pixels were used. In particular, visualization confirmed that YOLO v8 is excellent at detecting small objects. This experiment confirmed that finding a suitable YOLO model through learning experiments according to training data is essential. In the future, if we can obtain accumulated results for various objects through various training data, it is expected that it will be possible to develop a solution that can propose a YOLO version and hyperparameters that consider object characteristics.

Keywords : YOLO, VisDrone, Object Detection, Hyperparameter

### 초 록

객체 탐지는 자율주행, 안전관리, 도시관리, 스마트 팩토리, 환경보호 등 사회의 다양한 분야 활발하게 사용되고 있다. 한편, 다양한 버전의 YOLO가 발표됨에 따라 학습데이터에 적합한 모델 및 하이퍼 파라미터 선정에 관한 연구가 필요하다. 이에, 본 연구에서는 VisDrone-DET 학습데이터를 사용하여 다양한 버전의 YOLO(v5, v6, v8, v10 및 v11)의 성능을 학습 실험을 통해 확인하였다. 또한, 하이퍼 파라미터 중 가중치 크기, 입력 이미지 해상도, 최적화 기능 및 학습률의 변경을 통해 정확도의 변화도 분석하였다. 실험 결과, YOLO v8 XLarge 모델에 대해 Adam 최적화 함수, 학습률 0.001, 입력 사진 크기 1920픽셀을 사용하였을 때, mAP@50 기준 0.669로 정확도가 가장 높은 것으로 나타났다. 특히, 시각화를 통해 YOLO v8이 작은 물체 감지에도 뛰어난 성능을 확인하였다. 본 실험을 통해 학습데이터에 따라 학습 실험을 통해 적합한 YOLO 모델을 찾는 것이 중요함을 확인하였으며, 추후 다양한 학습데이터를 통해 다양한 객체에 대해 누적된 결과를 얻을 수 있다면, 객체 특징을 고려한 YOLO 버전 및 하이퍼 파라미터를 제안할 수 있는 솔루션 개발도 가능할 것으로 기대된다.

핵심어 : YOLO, VisDrone, 객체 탐지, 하이퍼 파라미터

Received 2024.11.26, Revised 2024.12.09, Accepted 2024.12.26

1) Dept. of GIS Engineering, Namseoul University (E-mail: cjm5310@nsu.ac.kr)

2) Member, Dept. of Future & Smart Construction Research, KICT(Korea Institute of Civil Engineering and Building Technology) (E-mail: younj@kict.re.kr)

3) Corresponding Author, Member, Dept. of Drone and GIS Engineering, Namseoul University (E-mail: shyoo@nsu.ac.kr)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. 서론

하드웨어와 인공지능 등의 발전으로 인해 다양한 딥러닝 기반의 모델이 개발되고 있으며, 특히 상용화 수준의 객체 탐지 모델의 공개로 그 활용처가 늘어나고 있다(Gašparović *et al.*, 2023; Kim, *et al.*, 2022). 현재 객체 탐지는 자율주행, 안전관리, 도시관리, 스마트 팩토리, 환경보호 등 사회의 다양한 분야에서 활발하게 사용 중이며 중요성 또한 지속해서 증가하고 있다(Kim, Park, *et al.*, 2022; Kim *et al.*, 2023; Kno *et al.*, 2022; Lee *et al.*, 2022; Shin *et al.*, 2022). 객체 탐지 모델은 초기 2-stage 기반의 R-CNN 계열 모델이 발표되었으나, 현재는 학습의 최적화와 탐지 속도에 유리한 1-stage 기반의 YOLO (You Only Look Once)가 다양한 분야에서 주로 활용되고 있다.

YOLO(Redmon, 2016)는 탐지와 분류를 동시에 진행하는 최초의 1-stage 탐지 방식의 모델로써 빠른 탐지 속도가 강점이다. YOLO는 버전 3까지 동일한 연구자에 의해 발표되었으나, 이후 다양한 연구자들에 의해 성능이 개선된 버전이 발표되었고, 현재는 버전 11까지 발표되었다. Kim *et al.* (2021)은 무인항공기(UAV: Unmanned Aerial Vehicle)로 취득한 영상을 사용하여 여덟 종류의 건물을 YOLO v5로 탐지하는 연구를 수행한 결과, 69.38%의 정밀도(precision)와 86.5%의 재현율(recall)을 확보하였다고 주장하였다. Geum *et al.* (2024)은 YOLO v5와 YOLO v8로 실시간 화재 감지 모델을 개발하였고, 두 모델에 대해 mAP@50 지표로 정확도 비교를 수행한 결과, 각각 83.5%와 85.9%의 정확도를 가지는 것으로 확인되었다. Im *et al.* (2023)은 합성개구레이더(SAR: Synthetic Aperture Radar) 영상으로 선박을 탐지하기 위해 YOLO v5를 활용하였는데, mAP@50 지표를 기준으로 정확도가 99.5%인 것으로 나타났다. Bak *et al.* (2023)은 여러 조식동물들을 모니터링하기 위해 YOLO v7과 v8을 사용하였는데, YOLO v8이 v7에 비해 탐지 정확도와 탐지 속도가 더 뛰어나고 색상 왜곡에 더 강인한 것으로 확인되었다는 연구 결과를 발표한 바 있다.

선행 연구를 분석한 결과, 각각의 연구 목적을 달성하기 위해 다양한 버전의 YOLO가 사용된 것으로 확인되었다. 그러나 다수의 연구가 하나의 버전만을 사용하거나 YOLO의 개발자가 제안한 기본값을 그대로 사용한 것으로 확인되었다. 딥러닝 모델은 그 특성상 학습데이터가 바뀌거나 같은 모델을 사용하더라도 하이퍼 파라미터가 변경되면 알려진 성능과 달라질 가능성이 크다. YOLO도 딥러닝 기반의 객체 탐지 모델이기에, 이러한 특징에서 벗어날 수 없으므로, 이에 관한 확인과 검증이 필요하다.

본 연구에서는 특정 학습데이터를 사용하였을 때, YOLO의 버전에 따라 정확도에 변화가 있는지, 특히 최신 버전의 YOLO가 반드시 좋은 성능을 보이는데에 관한 검증을 수행하였다. 이를 위해 YOLO v5, v6, v8, v10 그리고 2024년 10월에 발표된 최신 버전인 v11까지 실험 대상에 모두 포함하였다. 또한, 하이퍼 파라미터의 변경에 따른 모델의 성능 변화를 검증하기 위해 모델 가중치의 크기, 입력 사진의 크기, 최적화 함수, 학습률을 각각 변경해가며 모델에 대한 정확도 평가를 수행하였다. 이에 따른 본 연구의 연구 흐름도는 Fig. 1과 같다.

YOLO의 경우, 새로운 모델을 발표할 때 모델의 성능 평가를 위해 MS COCO 학습데이터를 공통으로 사용하고 있다. 즉, MS COCO는 최신 버전일수록 정확도가 높음이 이미 증명되었다는 것을 의미하므로, 본 연구에서는 MS COCO가 아닌 VisDrone 학습데이터를 실험 데이터로 사용하였다. VisDrone은 사람, 차량, 이륜차 등 일상에서 쉽게 접할 수 있는 총 10가지 객체에 대해 구축된 객체 탐지 전용 학습데이터로(Zhu *et al.*, 2021), 주로 공간정보 분야에서 탐지하기 원하는 대상이 다수 포함되어 있기에 선정하였다.

실험 결과, YOLO 모델 중에서는 v8의 성능이 가장 우수한 것으로 확인되었으며, 모델과 관계없이 Adam 손실함수와 0.001의 학습률을 사용하였을 때 모델의 정확도가 가장 높은 것으로 확인되었다. 특히, 입력 사진의 크기가 크면 클수록 모델의 정확도가 향상하는 것으로 나타나, 객체 탐지의 정확도와 입력 사진의 크기가 높은 상관관계를 가짐을 실험을 통해 확인하였다.

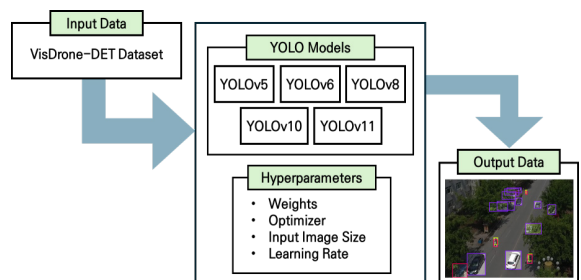


Fig. 1. Flow of study

## 2. YOLO 객체 탐지 모델 및 학습데이터

### 2.1 YOLO 모델의 버전별 특징

YOLO v5는 2020년 6월 ultralytics社에서 발표한 모델이다. YOLO v4까지는 C언어로 구현한 Darknet으로 모델을 개발하였으나 YOLO v5부터는 현재도 널리 사용되는 딥러닝 프

레이아웃 중 하나인 Pytorch로 개발하여 많은 사용자들이 편리하게 사용할 수 있고 객체 탐지 모델에 누구나 쉽게 접근할 수 있도록 개선하였다.

YOLO v5의 백본(backbone)은 YOLO v4에서 사용한 CSPDarknet53을, 헤드(head)는 YOLO v3와 v4에서 사용한 모듈을 사용하였다. 그러나 YOLO v5는 이전 모델에서 넥(neck)을 추가한 구조로 변경된 것이 가장 큰 특징이다. 넥의 역할은 다양한 크기의 객체 정보를 추출할 수 있도록 해상도별 이미지 피라미드를 만드는 것이다. 입력된 이미지를 다운 샘플링하여 서로 다른 크기의 다운 샘플링된 3개의 피쳐 맵을 생성한 뒤, 이를 업 샘플링하면서 서로 다른 크기의 새로운 피쳐 맵 3개를 생성하고 객체 정보를 추출한다. 이러한 구조로 인해 객체의 크기나 스케일에 상관없이 객체를 탐지할 수 있다는 강점이 있다.

YOLO v6는 2021년 10월 같은 개발사에서 발표한 YOLO v5를 개선한 모델이다. YOLO v5의 신경망보다 더 깊은 신경망을 구축하고 헤드의 탐지 스케일이 하나 추가된 등의 차이가 있지만, 기본적인 구조는 YOLO v5와 유사한 모델이다.

Ultralytics에서 개발하고 2023년 1월에 출시된 YOLO v8은 버전 5 및 6에 비해 여러 가지 아키텍처를 개선하였다. 기존에 사용하였던 백본의 C3 모듈은 성능은 유지하면서 매개변수의 개수를 줄일 수 있는 C2f 모듈로 교체되었고, 첫 번째 입력 컨벌루션 레이어의 필터 크기를  $6 \times 6$ 에서  $3 \times 3$ 으로 줄여 전체 매개변수 수가 감소시켰다. 이는 분류 정확도는 유지하면서 탐지 속도를 높이는 방안이기 때문에 적용된 것이다. 또한 각 업샘플링 프로세스가 수행되기 전에 배치되었던 두 개의 컨벌루션 레이어를 제거함으로써 모델을 더 단순화시켰으며, 박스 회귀와 분류 기능을 별도로 처리하는 분리된 헤드(Decoupled Head) 방식을 도입하여 속도와 정확도(AP) 향상을 도모하였다. 마지막으로 이전 버전에서 계속 사용하였던 객체성 분기를 제거하였는데, 이는 confidence score를 산출하는 데 사용하던 손실 중 하나로, 분류 정확도를 향상하기 위해 다른 손실 함수들을 도입하였다.

YOLO v10은 2024년 5월 Wang *et al.*(2024)이 발표한 모델이다. YOLO v8과 비교 시 가장 큰 차별점은 헤드에 있다. YOLO v10은 추론할 때 비최대억제(NMS: Non-Maximum Suppression)를 사용하지 않는다. 비최대억제의 특성상 수차례 계산을 통해 최선의 값을 추론하기 때문에 그만큼 추론 속도가 느릴 수밖에 없다. 따라서 YOLO v10은 비최대억제 대신 일대다, 일대일 헤드를 동시에 사용하여 추론하는 일관된 이중 할당(consistent dual assignments)을 구현하였다. 일대다 헤드가 여러 개의 예측 결과를 생성하면 일대일 헤드가

가 곧바로 최적의 추론 결과를 생성한다. 이로 인해 추론 시간을 단축하였고 비최대억제에 의존한 신경망의 비효율성을 해소하였다.

YOLO v11은 2024년 10월 ultralytics 사에서 발표한 모델로 현존 YOLO 모델 중 최신 모델이다. YOLO v11의 백본을 살펴보면 두 가지 변경점이 있다. 먼저 C2f 모듈이 C3k2로 교체되었다. C3k2는 하나의 커다란 컨벌루션 레이어 대신 이를 작게 수정한 두 개의 컨벌루션 레이어를 사용한 모듈이다. 또한 k2라는 명칭에서 볼 수 있듯 커널 크기가 이전 대비 더 작아졌다. 위와 같이 파라미터를 줄였음에도 성능이 저하되지 않았고 빠른 처리 속도를 구현하였다. 두 번째로 SPPF(Spatial Pyramid Pooling-Fast) 모듈 뒤에 C2PSA(Cross Stage Partial with Spatial Attention) 모듈을 추가하였다. C2PSA 모듈은 spatial attention을 강화하는 모듈로, 모델이 이미지의 주요 영역에 집중할 수 있도록 도와 탐지 정확도를 향상시켰다.

YOLO v11의 헤드에도 C3k2 모듈이 적용되었고, c3k 파라미터 사용 유무에 따라 모듈 적용 방법이 달라진다. c3k가 True이면 C3k2가 C3 모듈로 바뀌어 더욱 깊고 복잡한 객체 특징을 추출할 수 있고, c3k가 False이면 C3k2가 C2f와 유사하게 작동하여 특징을 추출한다. 이후 CBS 모듈을 거쳐 정확한 객체 탐지 능력을 갖추도록 헤드를 구성했다.

## 2.2 학습데이터

본 연구에서 사용한 VisDrone 데이터셋은 2019년에 발표된 객체 탐지용 오픈 데이터셋으로, 중국 14개 도시(톈진, 홍콩, 다칭, 간저우, 광저우, 진창, 류저우, 난징, 사오싱, 선양, 난양, 장자커우, 쑤저우, 쉬저우)의 도심과 교외에서 일상적으로 볼 수 있는 다양한 장면을 드론으로 촬영하여 구축하였다(DatasetNinja, 2024). VisDrone 데이터셋은 총 10,209 장의 정적 이미지와 263개의 동영상(179,264 프레임)으로 구성되어 있으며, 이미지 객체 탐지, 영상 객체 탐지, 단일 객체 추적, 다중 객체 추적 등의 네 가지 다른 목표에 활용할 수 있도록 데이터셋이 구분되어 있다.

본 연구에서는 VisDrone의 이미지 객체 탐지(DET) 데이터셋을 사용하였다. VisDrone-DET 데이터셋은 일상생활에서 자주 접할 수 있는 총 10종의 객체(보행자, 사람, 자동차, 밴, 버스, 트럭, 오토바이, 자전거, 차량이 있는 삼륜차, 삼륜차)에 대해 구축되어 있으며, 학습용과 시험용으로 각각 6,471장과 548장이 객체 탐지 전용으로 배포되고 있다. 본 연구에서는 인터넷에 공개된 학습데이터를 별도로 가공하지 않고 그대로 활용하였다. Fig. 2는 학습 데이터의 일부 사진을 나열한 것이다.





Fig. 2. Sample images from the VisDrone-DET (DatasetNinja, 2024)

### 3. 실험 결과 및 분석

#### 3.1 하이퍼 파라미터 설정

본 연구는 YOLO 버전과 하이퍼 파라미터 변경에 따른 분류 정확도의 변화를 확인하기 위해 가중치 크기, 입력 이미지 크기(image size), 최적화 함수(optimizer), 그리고 학습률(learning rate) 총 네 가지를 하이퍼 파라미터로 선정하였다. YOLO는 가중치의 크기에 따라 N(nano), S(small), M(medium), L(large), X(xlarge)로 나누어 가중치를 제공하기에, 이를 변경해가며 실험을 진행하였다. Nano 모델로 갈수록 정확도보다 빠른 탐지 속도가 중요한 상황이거나 저사양 하드웨어에서 사용할 수 있도록 적은 가중치로 설계한 것이고, XLarge 모델로 갈수록 속도보다 분류 정확도가 중요할 때 사용할 수 있도록 가중치의 개수를 많이 사용한 것이다. 따라서 분류 정확도 확보가 더 중요한 상황에서는 XLarge 모델을 사용하는 것이 더 좋다는 사실은 이미 알려진 것이나 VisDrone 데이터에 대해서도 동일하지 직접 확인하기 위해 가중치 크기를 하이퍼 파라미터로 선정하였다.

입력 이미지 크기의 경우, YOLO는 원본 이미지의 해상도와 관계없이 다양한 크기의 이미지를 입력받을 수 있도록 하나의 크기로 단일화하여 사용하는 기능 때문에 선정한 하이퍼 파라미터이다. 설정을 변경하지 않는다면 그 기본값은 640 픽셀로 설정되어 있는데, 이를 다시 말하면, 아무리 높은 해

상도의 이미지를 입력하더라도 결국 640 픽셀로 해상도를 감소시킨 뒤 사용한다는 의미이다. 이 경우, 원본 해상도에서는 탐지할 수 있는 객체를 탐지하지 못하고 놓치는 상황도 충분히 발생할 수 있다. 따라서 입력 이미지 크기를 증가시키면 정확도가 향상될 수 있을 것이라는 가설을 수립하게 되었다. 이에 본 연구에서는 일반적으로 카메라가 사진을 저장할 때 많이 사용되는 크기인 1280, 1360, 1600, 1920 픽셀로 입력 이미지 크기를 변경해가며 학습 실험을 수행하였으며, 1920을 마지막 수치로 선택한 이유는 VisDrone 학습데이터에서 제공하는 사진의 크기 중 가장 큰 사진의 크기가 1920 픽셀이기 때문이다.

최적화 함수는 손실 함수(loss function)의 최소값을 찾는 알고리즘을 말하는데, YOLO에서는 SGD (Stochastic Gradient Descent), Adam (Adaptive Moment Estimation), Adamax, AdamW, NAdam, RAdam, RMSProp 등 다양한 최적화 함수를 제공한다. 다양한 함수가 있지만, 본 연구에서는 딥러닝 연구에 가장 보편적으로 사용되며, 정확도 확보에 유리한 것으로 알려진 SGD와 Adam에 대해 학습 실험을 수행하였다. 추가적으로, YOLO를 개발한 Ultralytics에서는 SGD를 사용할 시 학습률을 0.01로, Adam을 사용할 시에는 0.001의 학습률 사용을 권장하고 있다. 이에, 본 연구에서는 해당 최적화 함수를 변경할 때마다 학습률도 함께 변경하였다.

### 3.2 정확도 평가 기법

딥러닝 모델의 정확도는 모델의 예측 결과와 실제 정답 간의 관계를 규정한 다양한 수식으로 측정할 수 있다. 본 연구에서는 YOLO 모델에서 정확도 평가에 대표적으로 사용하는 정밀도(precision), 재현율(recall), mAP(mean average precision)@50, mAP@50-95를 사용하였다. 정밀도는 모델이 정답으로 분류한 것 중 실제 정답인 것의 비율을 의미하며 (Eq. (1)), 재현율은 실제 정답 중 모델이 정답으로 분류한 것의 비율을 의미한다(Eq. (2)).

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

where  $TP$  represents the True Positive,  $FP$  the False Positive, and  $FN$  the False Negative.

mAP는 객체별로 산출된 AP의 평균으로, AP는 정밀도와 재현율 간의 관계를 PR 곡선(Precision-Recall curve)으로 그린 다음, 그 하부 면적을 의미한다. 여기서, @50과 @50-95의 차이는 IoU (Intersection Over Union)의 수치 차이로, mAP@50은 IoU가 0.5 이상인 객체를 대상으로 산출의 AP의 평균을 의미하며, mAP@50-95는 IoU 0.5 부터 0.95 이상까지 0.05씩 증가하며 산출한 AP의 평균을 의미한다.

### 3.3 실험 결과 및 분석

본 연구에서는 VisDrone-DET 학습데이터에 대해 YOLO 모델별 분류 정확도의 변화를 확인하기 위해 모델의 가중치 크기, 초기 입력 이미지의 크기, 그리고 최적화 함수를 변경하였으며, 각 실험 결과는 Table 1에 정리하였다. 첫 번째 가중치 크기의 경우, YOLO를 개발한 Ultralytics에서 제시하고 또 통상적으로 알려진 바와 같이 XLarge 모델의 분류 정확도가 Nano 모델보다 더 큰 것으로 확인되었다. YOLO v5에 대해 mAP@50 기준으로 XLarge 모델이 0.428, Nano 모델이 0.280으로 0.148만큼 큰 폭의 정확도 향상이 있는 것으로 나타났다기 때문이다.

두 번째는 최적화 함수와 학습률에 따른 정확도 평가 결과이다. YOLO v5, v8, v10, v11에 대해 SGD와 학습률 0.01, Adam과 학습률 0.001을 사용하였을 때 산출된 정확도를 비교한 결과, 모두 Adam과 학습률 0.001을 사용하는 것이 더 좋은 성능을 보이는 것으로 나타났다. YOLO v5, v6, v10의 경우, mAP@50 기준으로 입력 사진의 크기가 1280일 때, SGD

를 사용하였을 때보다 Adam을 사용하였을 때 더 나은 성능을 보였다. 한편, YOLO v8과 v11은 입력 사진의 크기가 1280일 때 SGD가 Adam보다 더 나은 성능을 가지는 것으로 나타났다, 최종적으로 입력 사진의 크기가 1920일 때에 대해서는 Adam의 성능이 더 나은 것으로 나타났다.

세 번째는 입력 이미지 크기에 따른 정확도 평가 결과이다. YOLO v5, v6, v8의 경우, 이미지 크기와 탐지 정확도가 정비례 관계임을 확인할 수 있었다(Fig. 3). Adam을 사용한 환경에서 mAP@50 기준 입력 이미지 크기 1280, 1360, 1600, 1920에 대해 YOLO v5는 각각 0.603, 0.606, 0.627, 0.641로, YOLO v6는 0.568에서 0.592, 0.610, 0.629로 점차 증가하였고, 특히 YOLO v8은 0.580, 0.639, 0.657, 0.669로 정확도가 큰 폭으로 증가한 것을 확인할 수 있었다. 한편, YOLO v10과 v11의 입력 사진 크기에 따라 점차적으로 정확도가 향상하지는 않았으나, 1280과 1920에 mAP@50 기준으로 YOLO v10이 각각 0.633에서 0.657로 YOLO v11은 각각 0.572, 0.647로 향상하는 것으로 확인되었다. 결국, 사진 입력 크기가 클수록 높은 분류 정확도를 가진다고 말할 수 있다고 판단된다.

마지막으로 YOLO 모델에 따른 정확도 평가 결과이다. Adam 최적화 함수를 사용하고, 입력 사진 크기는 1920을 사용하였을 때, YOLO v5, v6, v8, v10, v11이 mAP@50 기준으로 각각 0.641, 0.629, 0.669, 0.657, 0.647로, YOLO v8의 정확도가 가장 높은 것으로 나타났다. 참고로 본 실험을 통해 산출된 YOLO v8의 mAP@50 수치는 VisDrone-DET 학습데이터를 활용한 정확도 비교 평가 대회의 리더보드에 공개된 PP-YOLOE-plus 모델의 정확도인 0.667보다도 더 높은 수치이다(PapersWithCode, 2024).

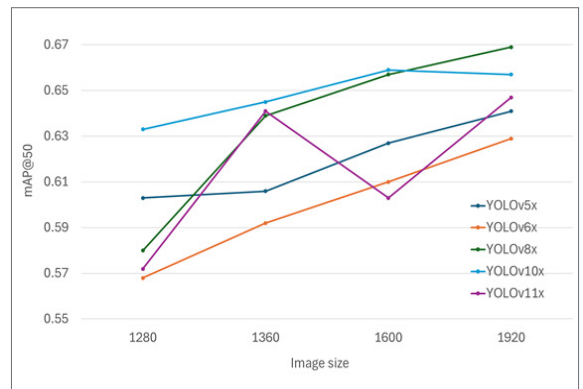


Fig. 3. Evaluation of accuracy for YOLO versions(Adam) based on input size

Table 1. Training accuracy by YOLO version (The highest accuracies are indicated in bold)

Version of YOLO	Optimizer	Models	Input image Size	Precision	Recall	mAP@50	mAP@50-95
PP-YOLOE-plus				-	-	0.667	-
YOLO v5	SGD	YOLOv5n	640	0.393	0.284	0.280	0.146
		YOLOv5x	640	0.537	0.424	0.428	0.259
	Adam	Scratch	1280	0.641	0.544	0.580	0.369
		YOLOv5x	640	0.519	0.417	0.421	0.251
			1280	0.665	0.565	0.603	0.389
			1360	0.663	0.569	0.606	0.393
			1600	0.673	0.594	0.627	0.408
			1920	0.695	0.603	0.641	0.422
YOLO v6	Adam	YOLOv6x	1280	0.623	0.542	0.568	0.359
			1360	0.658	0.554	0.592	0.374
			1600	0.657	0.576	0.610	0.392
			1920	0.674	0.595	0.629	0.414
YOLO v8	SGD	YOLOv8x	1280	0.656	0.555	0.588	0.377
			1920	0.699	0.623	0.658	0.435
	Adam	YOLOv8x	1280	0.648	0.547	0.580	0.370
			1360	0.680	0.611	0.639	0.417
			1600	0.690	0.623	0.657	0.432
			<b>1920</b>	<b>0.698</b>	<b>0.641</b>	<b>0.669</b>	<b>0.443</b>
YOLO v10	SGD	YOLOv10x	1280	0.674	0.592	0.618	0.402
	Adam	YOLOv10x	1280	0.668	0.610	0.633	0.413
			1360	0.695	0.606	0.645	0.424
			1600	0.708	0.619	0.659	0.434
			1920	0.688	0.618	0.657	0.435
YOLO v11	SGD	YOLOv11x	1280	0.660	0.575	0.600	0.387
			1920	0.677	0.600	0.633	0.418
	Adam	YOLOv11x	1280	0.636	0.541	0.572	0.365
			1360	0.674	0.611	0.641	0.419
			1600	0.664	0.565	0.603	0.389
			1920	0.669	0.621	0.647	0.428

Figs. 4, 5는 학습데이터에 포함된 사진에 대한 예측 결과를 모델별로 시각화한 것으로, 객체 차이를 명확히 확인하기 위해 입력 사진의 크기는 1280 픽셀과 1920 픽셀에 대해서만 나열하였다. YOLO의 버전별로 또 입력 사진의 크기별로 탐지된 객체의 개수는 달랐으나, YOLO v6와 v10을 제외하면 입

력 사진 크기를 크게 설정하였을 때, 더 많은 객체가 탐지되는 것으로 확인되었다. 특히, Figs. 4, 5에서 공통적으로 확인할 수 있는 점은 YOLO v8의 객체 탐지율로, 사람과 같이 크기가 작은 객체에 대한 탐지율이 다른 모델들에 비해 월등히 높은 것으로 확인되었다.



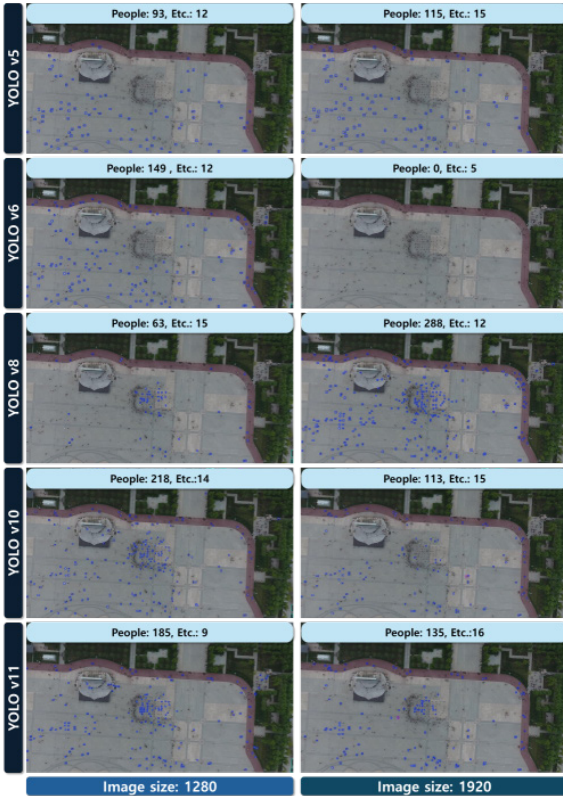


Fig. 4. Prediction results by YOLO version and input image size(0000059\_01886\_d\_0000114.jpg)

#### 4. 결론

본 연구에서는 객체 탐지 연구를 위해 공개된 VisDrone-DET 학습데이터를 사용하여 YOLO v5, v6, v8, v10, v11에 대해 학습 실험을 진행하였다. VisDrone-DET 학습데이터에 대해 버전별 정확도의 변화를 관찰하였으며, 모델 가중치의 크기, 입력 사진의 크기, 최적화 함수, 학습률을 변경해가며 하이퍼 파라미터의 변경에 따른 정확도 변화도 함께 확인하였다. 공개된 VisDrone-DET 학습데이터를 가공하지 않고 사용한 결과, YOLO v8 XLarge 모델에 대해 Adam 최적화 함수, 0.001의 학습률, 1920 픽셀의 입력 사진 크기를 사용하였을 때, mAP@50 기준 0.669로 가장 높은 정확도를 보였다. 또한, 시각적 분석을 통해 YOLO v8이 소형 객체 탐지에도 더 나은 성능을 보이는 것으로 확인되었다.

최신 버전의 YOLO가 더 나은 성능을 보이지 않은 결과에 대해 본 연구진은 YOLO 모델의 파라미터 개수가 영향을 미쳤을 것으로 판단하고 있다. YOLO는 정확도 향상보다는 탐

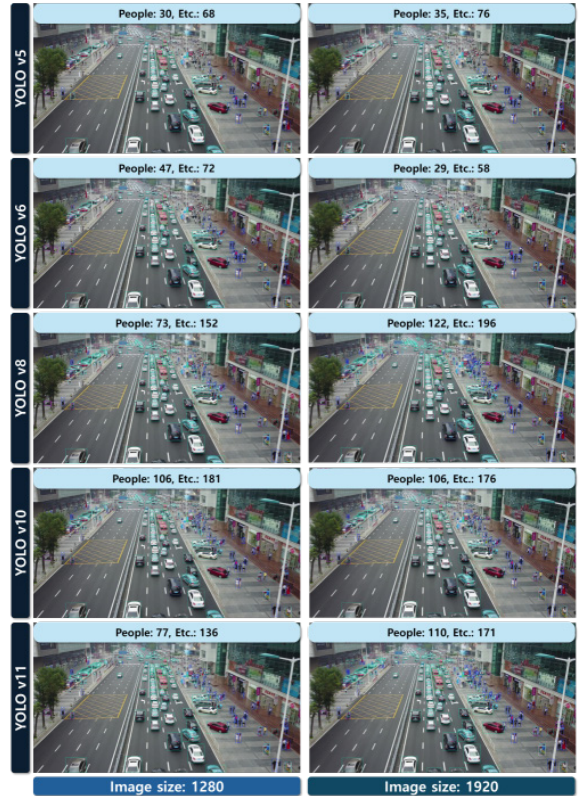


Fig.5. Prediction results by YOLO version and input image size(0000143\_01281\_d\_0000054.jpg)

지 속도 향상에 더 초점이 맞추어져 있기에, 되도록 가중치의 개수는 줄이면서 성능은 유지하는 것을 목표로 발전하고 있다. Ultralytics에서 공개한 YOLO v8 XLarge 모델의 가중치 개수는 68,229,648개이고, YOLO v11 XLarge 모델은 56,885,326개로, YOLO v11의 가중치 개수가 약 1천 1백만 개 더 적다. 결국, 이 만큼의 가중치가 줄어들면서 객체에 관한 정보를 잘 저장하지 못하여 발생한 현상으로 판단된다.

본 연구의 결과는 특정한 학습데이터만을 이용한 것이므로, 객체의 특징 또는 사진 취득 방식에 따라 상이한 결과가 도출될 수 있다. 본 연구에서 수행한 방법론을 다양한 학습데이터에 대해 적용한다면, 객체별 특징에 따라 적합한 버전의 YOLO와 하이퍼 파라미터를 제안할 수 있는 솔루션의 개발이 가능할 것으로 기대된다.

#### 감사의 글

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으

로 수행되었음(과제번호 : RS-2022-00143782, 차세대 디지털 국토정보 구축을 위한 고정/이동플랫폼 기반 동적주제도 구축 기술 개발(3/5)).

## References

- Bak, S., Kim, H.-M., Kim, T.-Y., Lim, J.-Y., and Jang, S.W. (2023), Evaluation of robustness of deep learning-based object detection models for invertebrate grazers detection and monitoring, *Korean Journal of Remote Sensing*, Vol. 39, No. 3, pp. 297-309. (in Korean with English abstract) <https://doi.org/10.7780/kjrs.2023.39.3.4>
- DatasetNinja. (2024), VisDrone2019-DET Dataset, *Dataset Ninja*, <https://datasetninja.com/vis-drone-2019-det> (last date accessed: 1 November 2024).
- Gašparović, B., Mauša, G., Rukavina, J., and Lerga, J. (2023), Evaluating YOLOv5, YOLOv6, YOLOv7, and YOLOv8 in underwater environment: Is there real improvement?, *2023 8th International Conference on Smart and Sustainable Technologies (SpliTech)*, pp. 1-4.
- Geum, K., Park, J., Lee, C., Choi, H., Go, Y., and Kim, d. (2024), Fire detection and notification system using YOLO object detection technology, *Proceedings of KIIT Conference*, pp. 267-270. (in Korean with English abstract)
- Im, Y., Youn, Y., Kang, J., Kim, S., Jeong, Y., Choi, S., Seo, Y., and Lee, Y. (2023), Ship detection from SAR images using YOLO: Model constructions and accuracy characteristics according to polarization, *Korean Journal of Remote Sensing*, Vol. 39, No. 5, pp. 997-1008. (in Korean with English abstract) <https://doi.org/10.7780/kjrs.2023.39.5.3.9>
- Kim, B.-R., Park, M.-S., Kim, J.-W., Do, Y.-B., Oh, S.-Y., and Yoon, H.-J. (2022), Analysis temporal variations marine debris by using Raspberry Pi and YOLOv5, *The Journal of the Korea Institute of Electronic Communication Sciences*, Vol. 17, No. 6, pp. 1249-1258. (in Korean with English abstract) <http://dx.doi.org/10.13067/JKIECS.2022.17.6.1249>
- Kim, J.-H., Kim, N., Park, Y.W., and Won, C.S. (2022), Object detection and classification based on YOLO-V5 with improved maritime dataset, *Journal of Marine Science and Engineering*, Vol. 10, No. 3, 377p. <https://doi.org/10.3390/jmse10030377>
- Kim, J., Lee, T., and Yeom, J. (2023), Personal mobility detection through application of YOLO deep learning algorithm to drone images, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol. 41, No. 4, pp. 239-246. (in Korean with English abstract) <https://doi.org/10.7848/ksgpc.2023.41.4.239>
- Kim, J.S. and Hong, I.Y. (2021), Analysis of building object detection based on the YOLO neural network using UAV images, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol. 39, No. 6, pp. 381-392. (in Korean with English abstract) <https://doi.org/10.7848/ksgpc.2021.39.6.381>
- Kno, Y., Jeong, K., Kweon, J., Cho, I., and Ju, Y. (2022), Development of smart factory abnormal detection method using object detection and AutoEncoder, *Global Business Administration Review*, Vol. 19, No. 5, pp. 76-90. (in Korean with English abstract) <https://doi.org/10.38115/asgba.2022.19.5.76>
- Lee, P. and Song, J.-W. (2022), Forest Fire Object Detection Using Deep Learning-Based Algorithm, *Journal of Digital Contents Society*, Vol. 23, No. 9, pp. 1869-1877. <http://dx.doi.org/10.9728/dcs.2022.23.9.1869>
- PapersWithCode. (2024), Object Detection on VisDrone-DET2019, *Papers with code*, <https://paperswithcode.com/sota/object-detection-on-visdrone-det2019-1> (last date accessed: 1 November 2024 ).
- Redmon, J. (2016) You only look once: Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- Shin, Y. and Yoo, Y. (2022), Smart factory safety system using artificial intelligence object detection, *KIIE Transactions on Computing Practices*, Vol. 28, No. 10, pp. 493-504. (in Korean with English abstract) <https://doi.org/10.5626/KTCP.2022.28.10.493>
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., and Ding, G. (2024), YOLOv10: Real-time end-to-end object detection, *arXiv preprint*, Vol. 2405, No. 14458. <https://doi.org/10.48550/arXiv.2405.14458>
- Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., and Ling, H. (2021), Detection and tracking meet drones challenge, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 44, No. 11, pp. 7380-7399. <https://doi.org/10.1109/TPAMI.2021.3119563>