

DETR를 이용한 작물 객체 인식 알고리즘 연구

정우혁, 이명훈*

*국립순천대학교

juh0329@naver.com, *leemh777@scnu.ac.kr

Research on Crop Object Detection Using the DETR Algorithm

Jung U Hyeok , Lee Meong Hun*

*Sunchon National Univ.

요약

본 연구는 스마트팜 환경에서의 작물 자동 인식을 위해 대표적인 객체 탐지 알고리즘인 YOLO(You Only Look Once)와 DETR(DEtection TRansformer)의 성능을 비교·분석하였다. 동일한 조건에서 촬영된 작물 이미지 데이터셋을 기반으로 두 모델을 학습시킨 후, 정확도, 평균 정밀도, 처리 속도 등 다양한 평가 지표를 활용하여 정량적 성능을 평가하였다. 실험 결과, YOLO는 상대적으로 빠른 추론 속도를 보였으나, DETR은 복잡한 배경과 부분 가림 상황에서도 더 높은 탐지 정확도를 나타냈다. 특히 DETR은 트랜스포머 기반의 글로벌 어텐션 메커니즘을 통해 작물의 위치와 형태를 더 정밀하게 인식할 수 있었다. 본 연구는 DETR이 실시간성이 요구되지 않는 정밀 농업 환경에서보다 효과적인 작물 인식 도구가 될 수 있음을 시사하며, 향후 실제 농가 환경에 적용한 확장 연구의 필요성을 제안한다.

I. 서론

농업 분야는 생산성 향상과 노동력 절감, 환경 부담 최소화를 동시에 달성하기 위한 디지털 전환의 흐름 속에 있다. 특히 스마트팜은 정보통신기술(ICT)을 농업에 접목하여 작물의 생육환경을 자동으로 제어하고 데이터를 기반으로 한 정밀농업을 실현하는 핵심 기술로 주목받고 있다[1].

특히, 카메라 기반의 객체 탐지 기술은 작물의 종류를 분류하거나 병해충 여부를 조기에 파악하는 데 활용되며, 노동력 절감과 생산성 향상에 기여할 수 있다. 이러한 배경에서 객체 탐지 알고리즘은 스마트팜 환경에서 핵심적인 요소로 주목받고 있다.

YOLO(You Only Look Once) 계열 모델은 빠른 처리 속도와 높은 실시간성으로 인해 다양한 실무 현장에서 널리 사용되어 왔다.[2] 반면, 최근에는 트랜스포머 기반의 DETR(DEtection TRansformer) 모델이 등장하면서 객체 탐지 분야에서 새로운 패러다임을 제시하고 있다. DETR은 글로벌 어텐션 메커니즘을 활용하여 복잡한 배경이나 겹침 상황에서도 정밀한 객체 탐지가 가능하다는 장점을 가지고 있다.[3]

그러나 농업 분야, 특히 작물 인식 분야에서 YOLO와 DETR의 성능을 직접 비교한 연구는 아직 부족한 실정이다. 작물 인식은 일반적인 객체 탐지보다 더 다양한 환경 변수에 영향을 받기 때문에 알고리즘 선택이 성능에 큰 차이를 유발할 수 있다. 이에 따라, 두 알고리즘의 성능 차이를 실험적으로 분석하고, 농업 분야에 더욱 적합한 모델을 제시하는 것이 중요하다.

본 연구에서는 동일한 작물 이미지 데이터셋을 활용하여 YOLO와 DETR의 성능을 비교·분석하고, DETR이 작물 인식에서 갖는 장점과 한계를 평가하고자 한다. 이를 통해 스마트팜 환경에 최적화된 객체 탐지 기술 선택에 기초 자료를 제공하는 것을 목적으로 한다.

II. 관련연구

YOLO 계열은 대표적인 One-stage 객체 탐지 모델로, 빠른 추론 속도와 비교적 간단한 구조 덕분에 실시간 응용 분야에서 널리 사용됐다. 특히

YOLOv5는 경량화된 구조와 학습 효율성으로 높은 활용성을 보이며 농업 분야에서도 많이 적용되고 있다.[4] 그러나 YOLO는 고정된 그리드 셀을 기반으로 객체의 위치를 예측하는 구조적 특성 때문에, 복잡한 배경이나 객체 간의 겹침이 많은 환경에서는 탐지 정확도가 낮아질 수 있으며, 형태가 불규칙하거나 경계가 모호한 객체에 대해서는 인식 성능이 저하되는 한계를 가진다. 또한 CNN 기반의 구조는 이미지 내의 장거리 의존 관계를 효과적으로 반영하기 어려워, 세밀한 맥락 정보 파악에 제약이 따른다.

이러한 한계를 극복하기 위해 제안된 DETR은 트랜스포머 기반의 구조를 객체 탐지에 도입함으로써, 전역적인 맥락 정보를 효과적으로 활용할 수 있는 새로운 접근 방식을 제시하였다.[5] DETR은 Anchor-free 및 NMS(non-maximum suppression)-free 구조로 설계되어 더욱 간결하고 해석 가능한 방식으로 객체 탐지를 수행하며, 복잡한 장면에서도 높은 탐지 성능을 보여준다. 특히 트랜스포머의 글로벌 어텐션 메커니즘은 객체 간의 관계를 학습하는 데 유리하여, 농업과 같이 다양한 배경 요소가 존재하는 환경에서 보다 안정적인 인식 결과를 제공할 수 있다.

III. 본론

본 연구에서는 동일한 작물 이미지 데이터셋을 활용하여 YOLO와 DETR의 객체 탐지 성능을 정량적으로 비교하였다. 두 모델 모두 동일한 학습 조건과 이미지 해상도에서 훈련되었으며, 테스트셋에 대해 오탐, 누락, 평균 정확도를 나타낼 수 있는 Precision, Recall, mAP(mean Average Precision), F1-score 지표를 기준으로 성능을 평가하였다.

표 1. 실험환경 및 학습설정

Table 1. Experimental environment and learning setup

항목	사양 및 설명
운영체제	Ubuntu 기반 워크스테이션
프레임워크	PyTorch (YOLO: Ultralytics / DETR: Facebook AI)
모델 버전	NVIDIA GPU
이미지 수집 장비	Intel RealSense D435i
해상도	1280 x 720
학습장비	NVIDIA RTX 3060 (12GB VRAM)
에폭 수	300 epochs
최적 알고리즘	Adam (lr=0.001)
증강 기법	Mosaic, HSV 변화, 좌우 반전 등
학습모델	YOLOv5, DETR

본 연구에서의 학습 및 실험은 표 1과 같이 동일한 환경에서 수행되었으며, 공정한 성능 비교를 위해 YOLO와 DETR 모두 같은 이미지 데이터셋, 전처리 조건, 하드웨어 사양을 적용하였다. 학습은 Ubuntu 기반의 워크스테이션에서 진행되었고, GPU를 활용하여 연산 효율을 극대화하였다. 수집된 총 2,500장의 이미지 중 1,800장은 학습에 사용되었고, 나머지 700장은 검증 및 테스트에 활용되었다.

YOLOv5는 PyTorch 프레임워크를 기반으로 하여 Ultralytics에서 제공하는 구조를 따랐으며, DETR은 Facebook AI에서 공개한 원본 구조를 바탕으로 학습하였다. 두 모델 모두 300 epoch 동안 학습하였고, Adam optimizer와 동일한 learning rate를 적용하였다.

그 결과, DETR은 모든 지표에서 YOLO보다 높은 성능을 나타냈다. YOLO는 실시간 처리를 고려한 경량화된 구조로 인해 비교적 빠른 추론 속도를 보였지만, 정밀도와 정확도 측면에서는 DETR에 미치지 못했다. 특히 mAP과 F1-score에서 DETR이 각각 0.84와 0.87을 기록하며, YOLO의 0.76, 0.80에 비해 뚜렷한 우위를 보였다. 이는 트랜스포머 기반의 글로벌 어텐션 메커니즘이 배경이 복잡하거나 작물이 겹쳐 있는 상황에서도 더 정교한 객체 인식이 가능하다는 점을 보여준다.

또한, DETR은 작은 객체에 대한 탐지에서도 상대적으로 우수한 성능을 보였다. 이는 CNN 기반 YOLO가 로컬 특징 중심으로 작동하는 반면, DETR은 이미지 전체의 컨텍스트를 활용하여 객체 간의 관계를 더 잘 이해하기 때문으로 판단된다.

아래의 그림 1의 그래프는 두 모델의 성능을 주요 지표별로 시각화한 것이다. DETR이 모든 항목에서 일관되게 높은 점수를 기록하고 있음을 확인할 수 있다

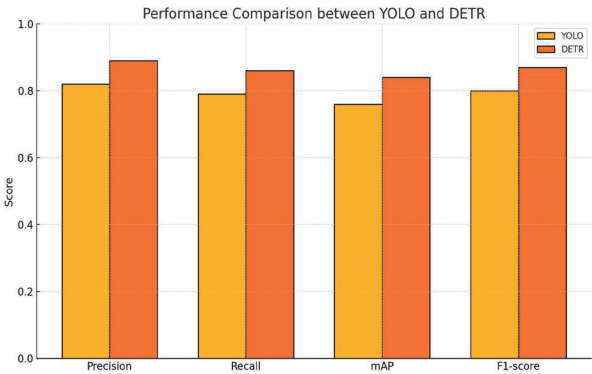


그림 1. YOLO와 DETR의 성능 비교

Figure 1. Performance Comparison Between YOLO And DETR

또한, YOLO는 다중 객체 탐지 기능을 갖추고 있어, 한 이미지 안에서 여러 개의 딸기 과일을 동시에 인식하고 추적할 수 있다는 점에서도 생육 데이터의 정량적 수집에 적합하다.

IV. 결론

작물 인식을 위한 객체 탐지 알고리즘으로 YOLO와 DETR의 성능을 동일한 조건에서 비교·분석한 결과 Precision, Recall, F1-score, mAP 등의 주요 성능 지표를 기준이 YOLO에 비해 DETR이 전반적으로 더 높은 정확도를 보이며, 복잡한 배경이나 객체 간의 겹침 상황에서도 안정적인 탐지 성능을 나타냈다.

YOLO는 빠른 추론 속도라는 장점을 가지지만, 그리드 기반 예측 구조의 한계로 인해 작은 객체나 불규칙한 형태의 작물을 인식하는 데에 있어 정확도에서 제한이 있었다. 반면, DETR은 트랜스포머 기반의 글로벌 어텐션 메커니즘을 통해 이미지 전역의 맥락을 고려한 정밀한 탐지가 가능하였으며, 특히 정확도 측면에서 현저한 우수성을 입증하였다.

따라서, 빠른 속도보다는 높은 정확도와 신뢰성이 요구되는 정밀 농업 및 스마트팜 환경에서는 YOLO보다는 DETR이 더 적합한 객체 탐지 모델임을 본 연구를 통해 확인할 수 있었다.

향후에는 DETR의 학습 속도 개선 및 경량화 모델을 적용하여 실제 농가 환경에서도 실용적으로 활용할 수 있도록 하는 추가 연구가 필요할 것이다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원을 받아 수행된 연구임 (IITP-2025-RS-2023-00259703, 기여율100)

참 고 문 헌

[1] 이승현, 김락우, 정득영, and 박대현, "ICT 기술을 적용한 농업분야의 연구 동향," 한국통신학회지(정보와통신), vol. 39, no. 10, pp. 3-12, 2022.

[2] 김삼근 and 안재근, "스마트 농업을 위한 YOLO 기반 작물 해충 탐지 모바일 애플리케이션," 한국산학기술학회 논문지, vol. 25, no. 7, pp. 603-610, 2024. (10.5762/KAIS.2024.25.7.603)

[3] 반충기, 박다영, 황영배, 이용, 장래영, and 최명석, "DETR 기반의 객체 임베딩을 활용한 이미지 검색," 한국차세대컴퓨팅학회 논문지, vol. 18, no. 5, pp. 84-92, 2022.

[4] 이승원, 유경현, 강영균, 김종민, and 이철수, "YOLO기반의 객체인식을 통한 딸기 수확 로봇 시스템에 대한 연구," 한국생산제조학회지, vol. 32, no. 2, pp. 101-108, 2023. (10.7735/ksmt.2023.32.2.101)

[5] 박예솔, 윤혜정, 김준오, and 장병탁, "학습되지 않은 객체 분할을 위한 트랜스포머 기반 RGB-D 융합," 한국정보과학회 학술발표논문집, pp. 420-422, 전남, 2024-12-18.