

YOLO 계열 모델을 이용한 KICE 및 시도교육청 수학 영역 기출문제 문제 영역 탐지 성능 비교

김수윤¹

¹부산중앙고등학교 3학년

star@devksy.xyz

Performance Comparison of YOLO Series Models for Problem Region Detection in Korean Mathematics Standardized Test Papers

Su-Yun Kim¹

¹Busan Jungang High School, Grade 12

요 약

본 연구는 한국교육과정평가원(KICE) 및 시도교육청에서 주최하는 수학 영역 기출문제지에서 문제 영역을 자동으로 탐지하기 위해 YOLO 계열 모델의 성능을 비교 분석하였다. 2024·2025년 기출문제 480문제를 대상으로 YOLOv8, YOLOv11, YOLOv12 모델을 동일한 조건에서 학습시켜 성능을 평가하였다. 실험 결과 세 모델 모두 mAP50 99.5% 수준의 우수한 성능을 보였으며, 학습 안정성 측면에서 YOLOv8, 최고 정확도 측면에서 YOLOv12가 적합한 것으로 나타났다.

1. 서론

한국교육과정평가원(KICE)이 주최하는 대학수학능력시험은 한국의 고등교육 진학을 결정짓는 중요한 평가로, 수험생 대부분은 KICE 및 시도교육청에서 제공하는 기출문제를 반복 학습하며 대비한다. [1] 그러나 기출문제는 주로 PDF 형태로 제공되어, 효율적인 학습 자료로 활용하기 위해서는 문제 영역을 체계적으로 가공하는 과정이 필수적이다.

이러한 문서 이미지에서의 문제 영역 분할 작업을 자동화하기 위해서는 객체 탐지 기법이 필요하며, 특히 실시간 처리가 가능한 딥러닝 기반 모델의 활용이 효과적이다. YOLO(You Only Look Once) 계열 모델은 이미지 내 객체를 실시간으로 탐지하는 대표적인 딥러닝 기반 기법으로, 단일 신경망을 통해 객체의 위치와 분류를 동시에 예측하여 높은 속도와 정확도를 달성한다는 장점이 있다. [2]

비교 기준점으로 널리 사용되는 YOLOv8은 다양한 응용 환경에서 안정적인 성능을 보이며, 객체 탐지 연구의 대표적 기준 모델로 활용되고 있다. [3] YOLOv11은 backbone 구조 최적화와 feature fusion 개선을 통해 연산 효율성을 향상시켰으며, YOLOv12는 기존 CNN 기반 아키텍처에서 벗어나 attenti

on 메커니즘을 도입한 attention-centric 설계로 모델링 능력을 크게 개선했다는 특징이 있다. [4] [5]

따라서 본 연구의 목적은 YOLOv8, YOLOv11, YOLOv12 세 모델을 동일한 실험 조건에서 학습시켜 KICE 수학 영역 기출문제지의 문제 영역 탐지 성능을 정량적으로 비교 평가하고, 각 모델의 성능을 분석하여 수학 문제 이미지 처리에 최적화된 모델을 제시하는 것이다.

2. 데이터 전처리

본 연구는 수학 영역 기출문제 중 2024년 3·5·6·7·9·10·11월과 2025년 3·6·9월 시행본, 총 10세트(각 48문항, 총 480문항)를 사용하였다. KICE 및 시도교육청에서 공개한 기출문제를 학술 연구 목적으로 활용하였다.

원본 PDF를 300 dpi PNG 형식으로 변환한 후, 각 문제 이미지를 대상으로 LabelImg 오픈소스 도구를 활용하여 문제 영역을 바운딩 박스로 라벨링하였다. [6] 라벨링된 데이터는 세트 단위로 분할하여 훈련용 7세트(336문항, 70%), 검증용 3세트(144문항, 30%)로 구성하였다. 추가로, 학습·검증에 포함되지 않은 과거 시험지(2023년 3·4월 시행본)를 독립 테스트셋으로 구성하였다.

30. 함수 $f(x)$ 를

$$f(x) = \lim_{n \rightarrow \infty} \frac{ax^{2n} + bx^{2n-1} + x}{x^{2n} + 2} \quad (a, b \text{는 양의 상수})$$

라 하자. 자연수 m 에 대하여 방정식 $f(x) = 2(x-1) + m$ 의 실근의 개수를 c_m 이라 할 때, $c_k = 5$ 인 자연수 k 가 존재한다.

$k + \sum_{m=1}^{\infty} (c_m - 1)$ 의 값을 구하시오. [4점]

(그림 1) 바운딩 박스

3. 모델 학습

모델 학습은 개인용 컴퓨터 환경(Windows 11, Intel i7-12700F CPU, NVIDIA GeForce RTX 4060 Ti 16GB GPU)에서 수행되었으며, PyTorch 기반 Ultralytics YOLO 라이브러리를 이용하였다. [7]

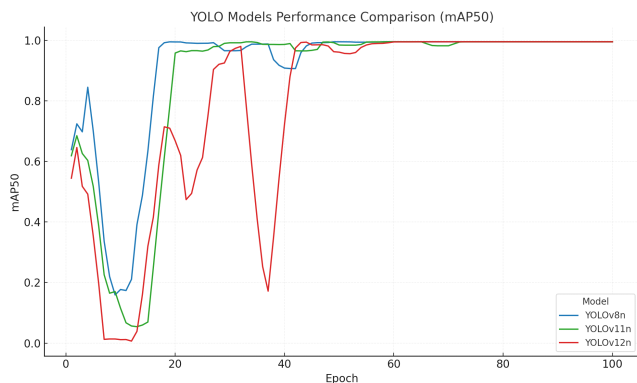
세 모델 모두 nano 버전을 사용하였으며, COCO 데이터셋으로 사전 학습된 가중치를 불러온 뒤, 기출문제지 데이터셋에 맞추어 파인튜닝(fine-tuning) 방식으로 학습하였다. 공정한 비교를 위해 모든 모델에 동일한 하이퍼파라미터를 적용하였으며, Ultralytics 공식 권장사항에 따라 batch size 16, epoch 100, learning rate 0.01 (cosine decay), 입력 이미지 크기 640×640을 사용하였다. [3]

4. 성능 평가

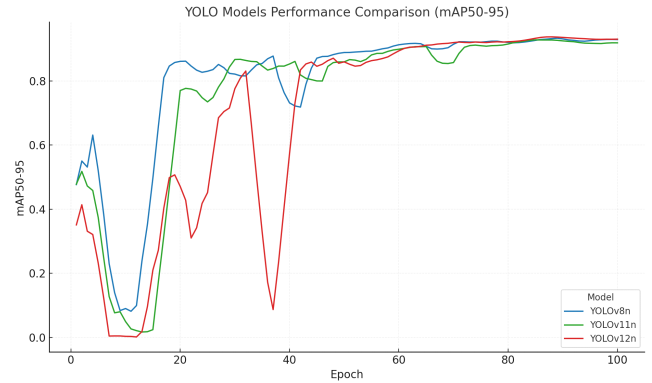
검증 성능이 가장 우수한 시점의 가중치(best.pt)를 평가에 활용하였다.

<표 1> 독립 테스트셋 기준 YOLO 모델별 성능 평가

Model	mAP50	mAP50-95	Precision	Recall
YOLOv8n	99.49%	91.81%	99.80%	100.00%
YOLOv11n	99.50%	91.93%	99.75%	100.00%
YOLOv12n	99.50%	92.59%	100%	99.00%



(그림 2) YOLO 모델별 mAP50 성능 변화 곡선



(그림 3) YOLO 모델별 mAP50-95 성능 변화 곡선

독립 테스트셋에서 모든 모델이 $mAP50 \geq 99.5\%$ 의 우수한 성능을 달성했으며, mAP50-95에서는 YOLOv12(92.59%) > YOLOv11(91.93%) > YOLOv8(91.81%) 순으로 나타났다. 그러나 학습 안정성 면에서 YOLOv8이 15 epoch에서 수렴한 반면, YOLOv12는 약 40 epoch가 필요했다.

5. 결론 및 제언

YOLOv12가 mAP50-95에서 92.59%로 가장 높은 성능을 기록했으나, 세 모델 간 최대 격차 0.78%p는 실용적으로 미미한 수준이었다. 학습 안정성과 수렴 속도를 고려하면 배포 기본 모델로 YOLOv8이, 최고 정확도가 필요하다면 YOLOv12가 타당하다. 모든 모델이 mAP50 99.5% 수준의 높은 탐지 성능을 보여 시험지의 규칙적 레이아웃이 탐지에 유리함을 시사한다. 향후 연구에서는 다양한 교과에 대한 일반화 성능 검증이 필요하다.

참고문헌

- [1] 한국교육과정평가원, "대학수학능력시험," <https://www.suneung.re.kr/>
- [2] J. Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," CVPR, 2016.
- [3] J. Terven, D. Cordova-Esparza, "The YOLO Framework: A Comprehensive Review," Computers, vol. 13, no. 12, 2024.
- [4] R. Khanam, M. Hussain, "YOLOv11: Key Architectural Enhancements," arXiv:2410.17725, 2024.
- [5] Y. Tian et al., "YOLOv12: Attention-Centric Real-Time Object Detectors," arXiv:2502.12524, 2025.
- [6] Tzutalin, "LabelImg," <https://github.com/HumanSignal/labelImg>
- [7] Ultralytics, "YOLO Documentation," <https://docs.ultralytics.com/>