유연 포장 제품 분할을 위한 파인튜닝 기반 YOLOv11과 zero-shot 모델의 플랫폼별 성능 비교

배성재, 김병준, 정성환, 서지환, 조세운*

*한국전자기술연구원

bsj940528@keti.re.kr, jun0420@keti.re.kr, shjeong@keti.re.kr, seojh410@keti.re.kr, *swcho@keti.re.kr

Performance Comparison of Fine-tuned YOLOv11 and Zero-shot Models for Flexible Package Segmentation Across Different Platforms

Bae Seong Jae, Kim Byoung Jun, Jeong Sung Hwan, Seo Ji Hwan, Cho Se Woon*

*Korea Electronics Technology Institute

요 약

본 논문은 식품 자동화 공정의 유연 포장 제품 검출을 위해 파인튜닝 기반 모델과 Zero-shot 모델의 하드웨어 플랫폼별 성능을 비교하였다. 파인튜닝 모델은 전 플랫폼에서 0.97 이상의 mAP를 기록하며 우수한 실시간 추론 성능을 입증했다. Zero-shot 모델은 별도의 추가 학습 없이 프롬프트만으로 추론 가능하나, 실시간성 및 정확도 면에서는 개선의 여지가 있음을 확인하였다.

Ⅰ. 서 론

최근 식품 자동화 공정에서는 유연 포장 제품과 같이 형상이 불규칙하고 변형이 잦은 객체를 안정적으로 검출하는 기술이 요구되고 있다. 기존의 파인튜닝 기반 객체 검출 및 분할 모델은 데이터 수집 및 라벨링 비용이 크고, 신규 제품 도입이나 포장 형태 변경 시 재학습이 필요하다는 한계를 가진다. 이에 따라 사전 학습된 대규모 모델을 활용하여 추가 학습 없이 객체 분할이 가능한 zero-shot 기반 분할 모델이 대안으로 주목받고 있으나, 실시간 처리 성능에 대한 검증이 필요하다.

본 논문에서는 유연 포장 제품 검출을 대상으로 YOLOv11 seg [1] 모델과 SAM(Segmentaion Anything Model) 계열의 zero-shot 기반 FastSAM [2] 및 SAM3 [3] 모델의 성능을 비교하고, RTX 4070 Ti, NVIDIA Jetson Orin, Jetson Thor 엣지 플랫폼에서의 추론 성능을 분석한다.

Ⅱ. 본론

2.1. 실험 환경

본 실험은 PC와 엣지 환경 간의 추론 성능 차이를 분석하기 위해 NVIDIA GeForce RTX 4070 Ti와 NVIDIA Jetson Orin, Jetson Thor 환경에서 수행하였다. 모든 실험은 동일한 테스트 데이터셋을 사용하였고 성능 평가는 mAP@0.5, Precision, Recall, F1-score와 같은 정확도 지표와 함께 FPS(Frame Per Second)를 기준으로 수행하였다.

YOLOv11과 FastSAM 모델 입력 해상도는 640×640으로 설정하였다. 이때 입력 이미지는 원본 종횡비를 유지한 상태에서 letterbox 방식으로 크기 조정되어 모델에 입력되었다. SAM3 모델은 고해상도 분할 품질을 고려하여 1280×720 해상도의 원본 이미지 입력을 사용하였다.

2.1.1. 하드웨어 플랫폼 구성

본 연구에서 사용한 하드웨어 플랫폼의 사양은 표 1에 정리하였다. RTX 4070 Ti는 고성능 데스크톱 환경을 대표하며, Jetson Orin과 Jetson Thor는 임베디드 AI 시스템에서의 실시간 추론 성능을 평가하기 위해 사용하였다. 특히 Jetson Thor는 차세대 고성능 임베디드 플랫폼으로, 대규모 비전 모델의 적용 가능성을 검토하기 위한 기준으로 포함하였다.

표 1. 하드웨어 플랫폼 성능 비교

플랫폼	PC	Jetson Orin	Jetson Thor
GPU SoC	NVIDIA GeForce RTX 4070 Ti	NVIDIA Jetson AGX Orin	NVIDIA Jetson Thor SoC
메모리	12 GB GDDR6X	64 GB LPDDR5	128 GB HBM
TFLOPS (FP16)	40.09	10.65	51.71

2.1.2. 비교 모델 개요

본 연구에서는 파인튜닝 기반 분할 모델과 zero-shot 기반 분할 모델의 특성을 비교하기 위해 YOLOv11-seg, FastSAM, SAM3 모델을 대상으로 실험을 수행하였다. 표 2는 각 모델의 파인튜닝 여부와 추론 시 프롬프트 사용 여부를 정리한 것이다.

YOLOv11-seg는 유연 포장 제품 데이터셋을 기반으로 파인튜닝 된 학습 기반 모델로, 별도의 프롬프트 없이 입력 영상만을 이용해 객체 분할을 수행한다. 반면 FastSAM과 SAM3는 추가적인 학습 없이 텍스트 프롬프트를 통해 객체를 분할하는 zero-shot 기반 모델로, 데이터 수집 및 라벨링 과정이 필요 없다는 장점을 가진다. 본 논문에서는 이러한 학습 방식 및 추론 구조의 차이가 GPU 플랫폼별 추론 성능과 실시간 처리 가능성에 미치는 영향을 비교해 분석한다.

표 2. YOLOv11, FastSAM, SAM3 모델 개요

모델	YOLOv11-seg	FastSAM	SAM3
파인튜닝	O	X	X
프롬프트	X	Text	Text

2.1.3. 테스트 데이터셋 및 모델 학습 설정

본 실험에 사용된 데이터셋은 유연 포장 제품을 대상으로 수집된 이미지

로 구성되었으며, 총 1,448장의 데이터를 학습, 검증, 테스트 세트로 분할하여 사용하였다. 구체적으로 학습 데이터는 1,231장, 검증 데이터는 136장, 테스트 데이터는 81장으로 구성하였다. 데이터 분할은 학습과 평가 간의 데이터 중복을 방지하기 위해 독립적으로 수행되었으며, 모든 성능 평가는 테스트 데이터셋을 기준으로 진행하였다.

표 3. 테스트 데이터셋

데이터 종류	개 수
학습 데이터	1,231
검증 데이터	136
테스트 데이터	81
총 데이터	1,448

YOLOv11-seg 모델은 지도 학습 기반 모델로, 사전 학습된(pretrained) 가중치를 초깃값으로 사용하여 학습을 수행하였다. 본 연구에서는 모델 구조에 따른 성능 차이를 분석하기 위해 YOLOv11n, YOLOv11s, YOLOv11m, YOLOv11l의 네 가지 모델 크기를 사용하였으며, 모든 모델은 동일한 학습 및 평가 설정을 적용하였으며, 이를 통해 GPU 플랫폼 및 모델 구조 차이에 따른 추론 성능 비교의 신뢰성을 확보하였다.

표 4. YOLOv11 모델 학습 파라미터

매개변수	값
모델	YOLOv11-seg
모델 종류	yolov11n-seg, yolov11s-seg, yolov11m-seg, yolov11l-seg
입력	640 * 640 이미지 (letterbox 적용)
출력	640 * 640 prediction mask
epoch	100

2.2. YOLOv11-seg 기반 파인튜닝 모델 실험 결과

그림 1은 YOLOv11-seg 모델을 다양한 GPU 플랫폼에서 평가한 추론 성능 결과를 나타낸다. YOLOv11-seg모델은 모든 GPU 플랫폼에서 mAP@0.5 기준 약 0.97 이상의 안정적인 검출 성능을 보였으며, 추론 속도 측면에서는 GPU 플랫폼에 따라 뚜렷한 차이가 관찰되었다. 데스크톱 환경(RTX 4070 Ti)에서는 YOLOv11n과 YOLOv11s 모델이 약 70 FPS 이상의 실시간 성능을 보였으며, Jetson Thor는 모든 모델에서 가장 높은 FPS를 기록했다.

Model	FPS (PC)	FPS (Thor)	FPS (Orin)	mAP@0.5	Precision	Recall	F1@0.5
YOLOv11n	73.77	164.64	50.48	0.978	0.968	0.951	0.959
YOLOv11s	72.42	152.66	42.57	0.978	0.978	0.924	0.950
YOLOv11m	61.63	94.76	32.47	0.977	0.973	0.950	0.961
YOLOv11l	48.99	73.40	25.29	0.979	0.977	0.931	0.953

그림 1. YOLOv11-seg 모델 성능 비교

2.3. Zero-shot 기반 SAM 계열 모델 실험 결과

그림 2는 zero-shot 기반 분할 모델인 FastSAM과 SAM3의 GPU 플랫폼별 추론 성능을 비교한 결과를 나타낸다. 본 실험에서 SAM3-Hiera-L 모델은 단일 텍스트 프롬프트인 “packet”을 사용하여 추론을 수행하였으며, FastSAM-s 모델은 유연 포장 제품의 다양한 외형을 포괄하기 위해 “packet”, “packaged food”, “plastic bag”, “wrapped package”의 다중 텍스트 프롬프트를 적용하였다.

FastSAM과 SAM3 모델은 YOLOv11 세그멘테이션 모델과 동일한 테스트 데이터셋을 기준으로 성능을 평가하였다. 실험 결과, SAM3는 mAP@0.5 기준 0.969의 높은 분할 정확도를 기록하여 학습 기반 모델과 유사한 수준의 분할 품질을 보였으나, FastSAM-s 모델은 모든 환경에서 40 FPS 이상의 속도를 기록하였으나, 다중 텍스트 프롬프트를 적용하였음에도 mAP@0.5는 0.665, Recall은 0.394로 상대적으로 낮은 분할 성능을 보였다.

Model	FPS (PC)	FPS (Thor)	FPS (Orin)	mAP@0.5	Precision	Recall	F1@0.5
FastSAM-s	173.88	167.86	41.68	0.665	0.665	0.394	0.495
SAM3-Hiera-L	24.11	11.34	5.64	0.969	0.969	0.896	0.931

그림 2. FastSAM, SAM3 모델 성능 비교

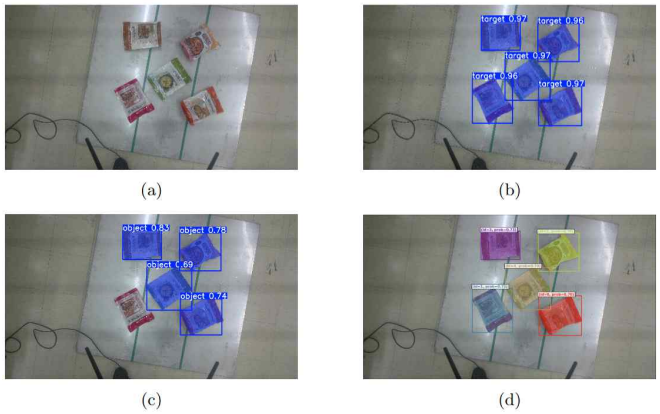

그림 3. 유연 포장 제품 모델별 분할 결과 비교: (a) 입력 이미지, (b) YOLOv11-seg 결과, (c) FastSAM 결과, (d) SAM3 결과

Ⅲ. 결론

본 논문에서는 유연 포장 제품 검출을 대상으로 학습 기반 YOLOv11-seg 모델과 zero-shot 기반 FastSAM 및 SAM3 모델의 성능을 엣지 및 PC GPU 환경에서 비교하고 분석하였다. 하드웨어 플랫폼 측면에서는 Thor 플랫폼이 여타 엣지 및 PC 환경 대비 가장 최적화된 연산 효율과 추론 속도를 기록하며 고속 공정 적용에 있어 최고의 성능을 입증하였다. YOLOv11 계열 모델은 모든 플랫폼에서 높은 분할 정확도와 실시간 처리 성능을 동시에 만족하여 산업용 자동화 공정에 가장 적합한 모델임을 확인하였다. zero-shot 기반 FastSAM과 SAM3 모델은 별도의 학습 없이 적용 가능하다는 장점을 가지지만, FastSAM은 분할 정확도 측면에서, SAM3는 실시간 처리 성능 측면에서 각각 한계를 보였다. 향후 TensorRT 기반 추론 최적화, 입력 해상도 조절, 프롬프트 엔지니어링 최적화 기법 등을 통해 zero-shot 기반 모델의 엣지 환경 적용 가능성 고도화에 관한 연구를 진행할 예정이다.

ACKNOWLEDGMENT

본 논문은 2025년도 산업통상자원부의 재원으로 한국산업기술기획평가원(KEIT) 로봇산업핵심기술개발사업 "다품종 소량 유연 포장 공정의 생산성 향상 위한 자동화 운영시스템 기술 개발(00508387)" 사업의 지원을 받아 수행된 연구결과임.

참 고 문 헌

[1] R. Khanam and M. Hussain, “YOLOv11: An Overview of the Key Architectural Enhancements,” arXiv preprint arXiv:2410.17725, 2024.

[2] H. Zhao, L. Zhang, X. Ding, et al., “Fast Segment Anything,” arXiv preprint arXiv:2306.12156, 2023.

[3] N. Ravi, E. Mintun, A. Kirillov, et al., “SAM 3: Segment Anything with Concepts,” arXiv preprint arXiv:2511.16719, 2025.