

# YOLO와 SAM을 활용한 해양 쓰레기 탐지 및 Pseudo Segmentation

윤서환<sup>1</sup>, 김상수<sup>1</sup>, 김지훈<sup>1</sup>, 송인섭<sup>1</sup>

<sup>1</sup>광운대학교 정보융합학부

## 1. 연구배경 및 목적

### 1.1 연구배경

해양환경 문제에 대한 국내외의 관심과 우려가 커지고 있는 가운데, 특히 해양 쓰레기가 우리의 생활환경과 생태계를 위협하는 주요 원인으로 지적되면서 세계적으로 중요한 환경 문제로 부상했다. 해양 쓰레기는 생태계뿐만 아니라 경제에도 심각한 영향을 미치며, 서식지 파괴, 종의 감소, 먹이사슬에 영향을 주어 생물 다양성을 감소시키는 등 부정적인 영향을 준다. 또한, 관광업과 수산업에도 경제적 손실을 초래한다.<sup>1)</sup>

기존의 해양 쓰레기 처리는 경제적, 환경적 손실을 줄이기 위해 매우 중요하지만, 넓은 해양 지역과 심해 환경에서 인간이 직접 이를 탐지하기란 어렵다.<sup>2)</sup> 이를 극복하기 위해서는 해양 쓰레기를 효과적으로 탐지하고 관리하기 위한 기술이 필요하다. 본 연구에서는 딥러닝 기술을 활용한 객체 탐지 모델을 사용하여 해양 쓰레기의 위치와 분포를 정밀하게 파악하고, 실시간 모니터링을 통해 빠른 조치를 가능하게 한다.

쓰레기 크기를 추정하기 위해서는<sup>3)</sup> segmentation이 필요하다. 그러나, segmentation을 위해서는 data labeling 작업이 필수적이며, 이 과정에서 상당한 비용이 발생한다. 예를 들어, 아마존의 데이터 라벨링 서비스는 bounding box label 생성에 이미지당 0.036\$, segmentation label 생성에는 약 0.84\$로 bounding box 비용의 약 23배가 필요하다.<sup>4)</sup> 구글 AI 플랫폼의 데이터 라벨링 서비스는 등급에 따라 bounding box 1,000개당 63\$, 49\$, segmentation label 1,000개당 870\$, 850\$로, segmentation label의 비용이 약 13배 정도<sup>5)</sup> 더 높다. 서비스와 데이터 종류에 따라 차이가 있지만, 전반적으로 segmentation label은 더 많은 비용이 든다. 본 연구에서는 해양 쓰레기 객체 탐지뿐만 아니라, 형태와 크기를 더 정확히 파악하여 효율적인 분류와 처리를 위한 segmentation까지 진행한다.

따라서, 해양 쓰레기를 효과적으로 탐지하고, 데이터 구축 비용 절감을 위한 연구가 필요하다. 본 연구에서는 객체 탐지에 효과적인 YOLO를 사용하여 해양 쓰레기를 효과적으로 탐지하고, 쓰레기의 크기를 추정하기 위해 SAM을 이용하여 YOLO를 통해 얻어진 bounding box를 SAM의 prompt로 사용하여 pseudo segmentation label을 얻고자 한다.

### 1.2. 연구목표

본 연구의 목표는 크게 두 가지다. 첫째, 환경 보호를 위해 해양 침적 쓰레기 이미지를 활용한 딥러닝 모델을 구축하여 해양 쓰레기 탐지의 정확도를 높이고, 자동적이고 효율적인 모니터링 기술에 기여한다. 둘째, Bounding box label만을 이용하여 더 상위 label인 pseudo segmentation label을 구축한다. 이를 통해 데이터 구축 비용을 절감하고, weakly supervised learning<sup>6)</sup>에 기여한다. 궁극적으로, 더 높은 수준의 label을 제공함으로써 해당 분야에 대한 심층적인 연구를 촉진할 수 있다.

## 2. 사용한 딥러닝 기술 소개

본 연구에서는 크게 YOLO, SAM, Sea-thru 3가지의 딥러닝 기술을 다루며, 각각의 기술에 대한 설명은 아래와 같다.

**YOLO.** YOLO<sup>7)</sup>는 "You Only Look Once"의 약자로, 실시간 객체 탐지(real-time object detection)의 대중화에 기여한 혁신적인 알고리즘이다. 2015년에 Joseph Redmon에 의해 처음 소개되었고, 그 이후로 여러 버전(v1~v10)이 출시되면서 계속해서 개선되고 있다. 본 연구에서는 v8 버전을 사용하며, 가장 작은 모델인 YOLOv8n<sup>8)</sup>을 사용했다.

**SAM.** SAM(Segment Anything)<sup>9)</sup>은 Meta에서 개발한 Foundation segmentation model로 prompt(point, bounding box, mask)를 이용하여 효과적인 segmentation을 달성했다. 본 연구에서는 YOLO를 사용하여 예측한 bounding box를 SAM의 prompt로 이용하여 segmentation을 진행했다. 이를 통해 pseudo label을 생성할 수 있다.

**Sea-thru.** Sea-thru<sup>10)</sup>는 2019년 CVPR에서 공개된 방법으로, 수중 이미지에서 물의 영향으로 발생하는 색상 왜곡 문제를 해결하여 원래의 색상으로 복원하는 기술이다. 일반적으로 사용되는 대기 중 빛의 전파를 설명하는 방식 대신, 수중 이미지의 특성을 반영해 거리와 반사율을 기반으로 색상을 보정하며, 빛의 산란 정도를 정확히 계산하여 더욱 정밀한 복원 성능을 제공한다. 그러나 해당 모델의 추론 시간이 오래 걸려 본 연구에서는 적용하지 않았다. 해당 기술에 대한 추가적인 자료는 별첨에 첨부한다.

## 3. 실험 설정 및 결과

### 3.1 실험 설정

**데이터셋.** AI-Hub에서 제공하는 재난 안전 데이터 중 해양 침적 쓰레기 이미지를 사용했다. 해당 데이터는 크게 소나 조사 이미지 약 46,000장과 수중 촬영 이미지 약 18,000장으로 구성되어 있으며, 본 연구에서는 수중 촬영 이미지만을 사용했다. 18,000장의 이미지 중 객체가 여러 개 존재하거나 존재하지 않는 이미지는 제거하여 약 12,000장의 이미지를 사용했다. 학습을 위해 train, validation, test의 비율을 7:1:2로 설정하였으며, 별도의 증강 없이 ImageNet normalization, resize만을 이용하여 전처리했다.

**Architecture.** 본 연구에서는 학습 모델로 YOLOv8n과 pseudo labeling을 위한 SAM 두 가지 모델을 사용했다. YOLOv8n의 경우, ultralytics에서 제공하는 모델을 사용했으며, COCO 데이터셋으로 사전 학습된 모델을 해당 데이터셋으로 전이 학습시켰다. optimizer로 AdamW(lr=0.001111, momentum=0.9)를 사용했으며, batch size는 128로 설정한 후 총 50 epoch을 진행했다. SAM의 경우 backbone으로 사전 학습된 ViT-H<sup>11)</sup>를 사용했으며, prompt는 YOLO를 통해 만들어진 bounding box를 사용하여 segmentation을 진행했다.

### 3.2 실험 결과

본 실험의 Metric은 class 예측값에 대한 precision, recall, f1 3가지와 bounding box에 대한 mAP50, mAP50-95에 대해 평가했다. 표 1을 보면, YOLOv8s의 결과가 train보다 test에서 더 높은 것으로 보아 모델이 과적합 되지 않고 일반화된 성능을 보여주고 있다.

또 AI-Hub의 기준 점수인 mAP50 0.75보다 더 높은 성능을 보여준다. 해당 실험에 대한 추가적인 자료는 별첨에 첨부한다.

표 1. YOLOv8n에 대한 train, test 결과

	precision	recall	f1	mAP50	mAP50-95
train	0.86	0.68	<b>0.74</b>	0.73	0.61
test	<b>0.91</b>	<b>0.69</b>	0.73	<b>0.76</b>	<b>0.62</b>

그림 1. SAM with Good segmentation

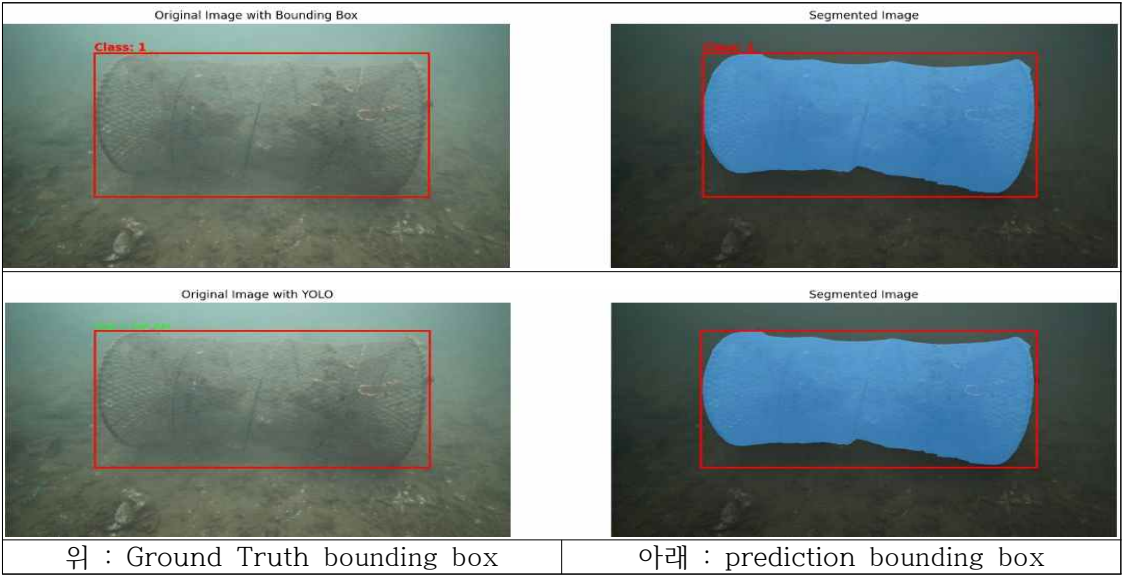
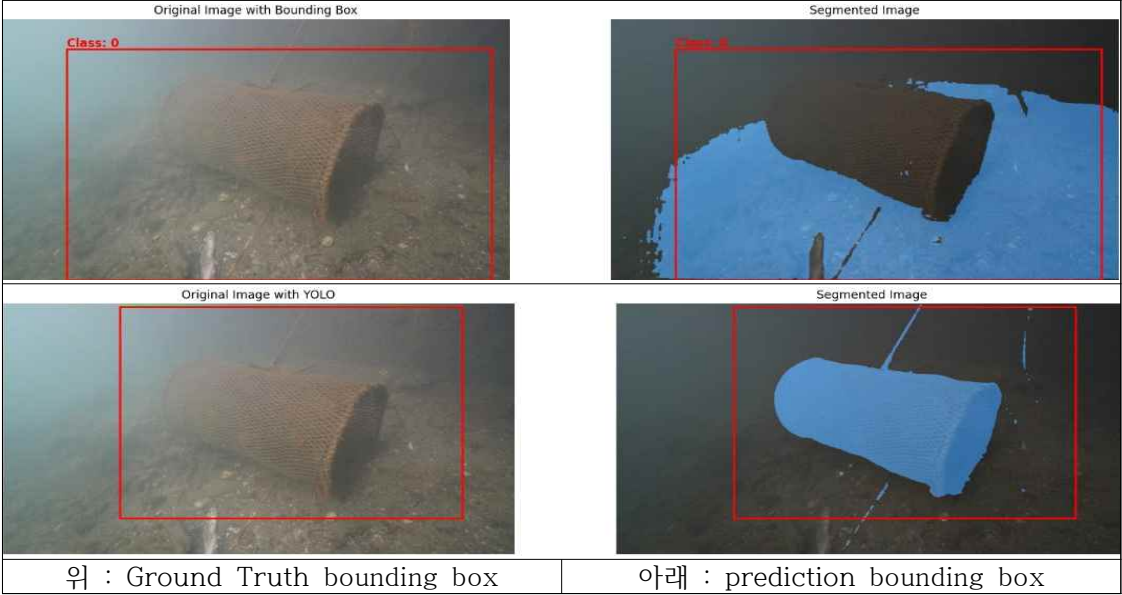


그림 2. SAM with Bad segmentation



SAM의 경우 Ground Truth, YOLO prediction을 통한 bounding box를 prompt로 사용하여 segmentation으로 진행했다. **그림 1**은 Ground Truth, YOLO 모두에서 좋은 segmentation을 보여줬다. 그러나 **그림 2**에서는 Ground Truth의 segmentation이 YOLO에서보다 부정확하게 인식되었다. 즉, 우리가 학습한 YOLO가 Ground Truth보다 강건한 bounding box를 제공하기 때문에 SAM에서도 강건한 성능을 보여준다. 게다가, SAM의 추론 시간은 평균 0.05초로, 실제 수십 분이 걸리는 작업 시간을<sup>12)</sup> 약 10,000배 가량 단축시켰다. 따라서 본 연구는 단순히 효과적인 해양 쓰레기 탐지에서 그치지 않고, pseudo label을 생성하는 비용과 시간을 수천 배 이상 절감시키는 효과를 가져온다.

## 4. 기대효과 및 활용분야

### 4.1 기대효과

**해양 환경 보호.** 본 연구에서는 해저 쓰레기를 탐지하는 YOLO와 탐지된 쓰레기를 자동으로 segmentation하는 SAM을 구축했다. 해당 두 모델을 통해 해양 침적 쓰레기를 신속하고 정확하게 탐지하고 분류할 수 있다. 이를 통해 해양 쓰레기를 조기에 발견하고 제거할 수 있어 해양 생태계를 유지하는 데 큰 도움이 될 것이다. 또 해양 쓰레기의 제거는 해양 생물의 생존을 보장하고, 생태계의 파괴를 막아 생물 다양성을 보호할 수 있다. 특히, segmentation을 통해 해양 쓰레기의 크기와 밀도를 추정할 수 있으며, 이는 쓰레기 처리의 우선순위를 정하고 수색 시간을 줄이는 데 기여한다.

**비용 절감.** 기존의 해양 쓰레기 탐지 방법은 많은 인력과 자원이 필요하다. 그러나 본 모델을 활용하면 자동화된 탐지와 탐지된 쓰레기의 segmentation을 동시에 수행할 수 있어 인력과 비용을 크게 절감할 수 있다. 이를 통해 해양 쓰레기 처리를 위한 비용 절감에 크게 기여할 수 있다.

게다가, SAM을 통해 해양 쓰레기에 대한 segmentation labeling 비용을 획기적으로 줄일 수 있다. 비용적인 측면에서는 약 23배, 시간적인 측면에서는 약 10,000배 정도로 기존 labeling 비용을 크게 줄일 수 있다. 특히, **그림 2**와 같이 본 모델은 Ground Truth보다 더 정확한 bounding box를 예측하고 더 좋은 label을 구축한다. 따라서 본 연구를 통해 해양 쓰레기를 처리하는 비용과 데이터 구축 비용을 크게 절감할 수 있다.

### 4.2 활용분야

**해양 쓰레기 탐지.** 국가의 공공기관 및 민간 기업에서 본 모델을 활용하여 해양 쓰레기의 탐지를 실시간으로 모니터링할 수 있다. 이를 통해 해양 환경 보호에 기여할 수 있으며, 더 나아가 해당 과정을 통해 축적된 데이터를 활용하여 더 고도화된 모델과 기술을 개발할 수 있다.

**해양 쓰레기 연구.** 해양 학자들과 연구자들은 본 모델을 활용하여 해양 쓰레기의 종류와 분포를 분석할 수 있다. 특히, pseudo segmentation label을 통해 쓰레기의 상태(변형, 오염 등)를 분석하는 연구를 수행할 수 있으며, 이는 수질 오염 원인과 해양 쓰레기 처리 방법을 결정하는데 도움을 줄 수 있다. 나아가, weakly supervised learning에도 기여할 수 있다. 이는, 하나의 연구분야가 더 발전될 수 있는 초석을 제공할 뿐만 아니라 보다 심층적인 연구를 촉진할 수 있다.

## 5. 역할 분담

윤서환 : 자료 조사, 데이터 전처리, 모델 학습, SAM

김상수 : 자료 조사, 데이터 수집, 모델 학습

김지훈 : 자료 조사, 데이터 수집, 데이터 전처리, 모델 학습

송인섭 : 자료 조사, Sea-thru, 모델 학습

- 
- 1) S. Werner, A. Budziak, J. Franeker, F. Galgani, G. Hanke, T. Maes, M. Matiddi, P. Nilsson, L. Oosterbaan, E. Priestland, R. Thompson, J. Veiga, and T. Vlachogianni, "Harm caused by Marine Litter," 2016.
  - 2) D. Papageorgiou, K. Topouzelis, G. Suaria, S. Aliani, and P. Corradi, "Sentinel-2 Detection of Floating Marine Litter Targets with Partial Spectral Unmixing and Spectral Comparison with Other Floating Materials (Plastic Litter Project 2021)," *Remote Sens.*, vol. 14, p. 5997, 2022.
  - 3) Kylili, Kyriaki, et al. "Identifying floating plastic marine debris using a deep learning approach." *Environmental Science and Pollution Research* 26 (2019): 17091-17099.
  - 4) "Amazon SageMaker Ground Truth pricing," Amazon Web Services, <https://aws.amazon.com/ko/sagemaker/groundtruth/pricing/>, accessed June 8, 2024.
  - 5) "AI Platform Data Labeling Service pricing", Google Cloud, <https://cloud.google.com/ai-platform/data-labeling/pricing?hl=ko>, accessed June 8, 2024.
  - 6) Zhou, Zhi-Hua. "A brief introduction to weakly supervised learning." *National science review* 5.1 (2018): 44-53.
  - 7) Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
  - 8) <https://github.com/ultralytics/ultralytics>
  - 9) Kirillov, Alexander, et al. "Segment anything." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.
  - 10) D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1682-1691.
  - 11) Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).
  - 12) <https://developer.nvidia.com/blog/segment-objects-without-masks-and-reduce-annotation-effort-using-the-discobox-dl-framework/>