



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

Segment Anything Model
(SAM)을 활용한 광학 및 SAR
위성 융합 영상 기반 객체 추출



HANSUNG
UNIVERSITY

2025년

한 성 대 학 교 대 학 원

융 합 보 안 학 과

융 합 보 안 전 공

이 하 영

석사학위논문
지도교수 이기원

Segment Anything Model
(SAM)을 활용한 광학 및 SAR
위성 융합 영상 기반 객체 추출

Object Extraction from Fused Optical and
SAR Satellite Imagery Using
the Segment Anything Model (SAM)



HANSUNG
UNIVERSITY

2025년 6월 일

한 성 대 학 교 대 학 원

융 합 보 안 학 과

융 합 보 안 전 공

이 하 영

석사학위논문
지도교수 이기원

Segment Anything Model
(SAM)을 활용한 광학 및 SAR
위성 융합 영상 기반 객체 추출

Object Extraction from Fused Optical and
SAR Satellite Imagery Using
the Segment Anything Model (SAM)

위 논문을 공학 석사학위 논문으로 제출함

2025년 6월 일

한 성 대 학 교 대 학 원

용 합 보 안 학 과

용 합 보 안 전 공

이 하 영

이하영의 공학 석사학위 논문을 인준함

2025년 6월 일



HANSONG
UNIVERSITY

심사위원장 석 병 진 (인)

심 사 위 원 서 화 정 (인)

심 사 위 원 이 기 원 (인)

국 문 초 록

Segment Anything Model (SAM)을 활용한 광학 및 SAR 위성 융합 영상 기반 객체 추출

한 성 대 학 교 대 학 원
융 합 보 안 학 과
융 합 보 안 전 공
이 하 영

2023년 4월, Meta AI는 제로샷(Zero-shot) 학습 기능을 지원하는 대규모 영상 분할 모델인 Segment Anything Model (SAM)을 발표하였다. 이후 객체 분할 기술이 다양한 분야에서 주목받기 시작했다. 특히, 원격탐사 분야에서 위성 영상을 활용한 수체 추출은 중요한 연구 주제로 수체의 정확한 식별은 환경 모니터링, 재해 관리, 수자원 평가, 도시 및 농업 계획 등 도움을 제공한다. 따라서 이번 연구에서는 고해상도 위성 영상인 KOMPSAT-3 광학 영상과 KOMPSAT-5 Synthetic Aperture Radar (SAR) 영상을 융합한 후 SAM을 활용하여 수체를 추출하였다. 일반적으로 광학 영상은 색상 정보를 제공하지만, 구름이나 대기 조건에 의해 식별 정확도가 저하될 수 있다. 반면 SAR 영상은 날씨 및 구름의 영향을 받지 않아 관측에 유리하지만, 노이즈가 많고 텍스처 정보가 부족하다. 이러한 문제를 해결하기 위해 두 영상의 특성을 결합하는 Wavelet 기반 이미지 융합 기법을 적용하였다. 융합 영상과 KOMPSAT-3 광학 영상을 사용하여 각각 SAM을 적용하였으며 평가를 위

해 국토지리정보원의 수계 수치지도와 Open Street Map (OSM)의 수체 자료를 비교 자료로 사용하였다. mean Intersection over Union (mIoU) 와 가중치 Weighted IoU (WIoU) 지표를 사용하여 값을 분석한 결과, SAR 영상을 융합한 영상에서 수체 탐지 정확도가 향상됨을 확인하였다. SAM의 세로썬 분할 기능은 일반적인 객체에는 강점을 보이나, 다양한 형태를 가진 수체와 같이 특화된 객체에 대해서는 성능이 제한적일 수 있다. 이를 보완하기 위해 GLH-Water 데이터셋을 활용하여 미세조정(Fine-tuning)을 수행하였다. 또한 실제 환경을 위해 Google Street View (GSV)로부터 국내 수부 이미지를 수집하여 데이터셋에 추가하였다. 기본 모델에 비해 Fine-tuned 모델이 복잡하고 다양한 형태의 수체를 보다 정확하게 추출하며 성능이 향상됨을 확인하였다. 이는 고해상도 위성 영상에서의 수체 탐지 성능이 영상 융합 기법과 SAM의 미세 조정을 통해 효과적으로 향상될 수 있음을 보여준다. 특히, 다양한 지형과 환경 조건에서도 경계가 복잡한 수체를 안정적으로 분할할 수 있다는 점에서, 재난 대응, 수자원 관리 등 공간 정보 기반의 다양한 실무 분야에 실질적인 활용 가능성을 제시한다.

【주요어】 KOMPSAT-3, KOMPSAT-5, Segment Anything Model, Water body, Wavelet-based Image Fusion, Fine-tuning

목 차

제 1 장 서 론	1
제 1 절 연구 배경	1
제 2 절 연구 목적	2
제 2 장 객체 분할 기술 및 연구 동향	4
제 1 절 객체 탐지 및 분할 기술	4
제 2 절 Segment Anything Model (SAM)	5
제 3 절 Geo-SAM	8
제 4 절 원격탐사 기반 객체 추출 연구	10
제 3 장 연구 자료 및 방법	13
제 1 절 연구 지역 및 연구 자료	13
제 2 절 연구 방법	15
제 3 절 이미지 융합 기법	16
1) 원격탐사 기반 이미지 융합 기법	16
2) 연구에서 사용한 Wavelet 기반 융합	18
제 4 절 Fine-tuning 및 데이터셋	20
1) Fine-tuning	20
2) GLH-Water 데이터셋	22
3) 학습 데이터셋 생성	25
4) 데이터셋을 이용한 Fine-tuning	29
제 5 절 수채 추출 실험	31
1) Geo-SAM 적용	31
2) SAM 적용	31

3) Fine-tuning 적용	32
제 4 장 수체 추출 결과 분석	36
제 1 절 비교 자료	36
제 2 절 성능 평가 지표	38
1) mean Intersection over Union (mIoU)	39
2) Weighted Intersection over Union (WIoU)	39
제 3 절 결과 분석	43
제 5 장 결 론	44
참 고 문 헌	46
ABSTRACT	52

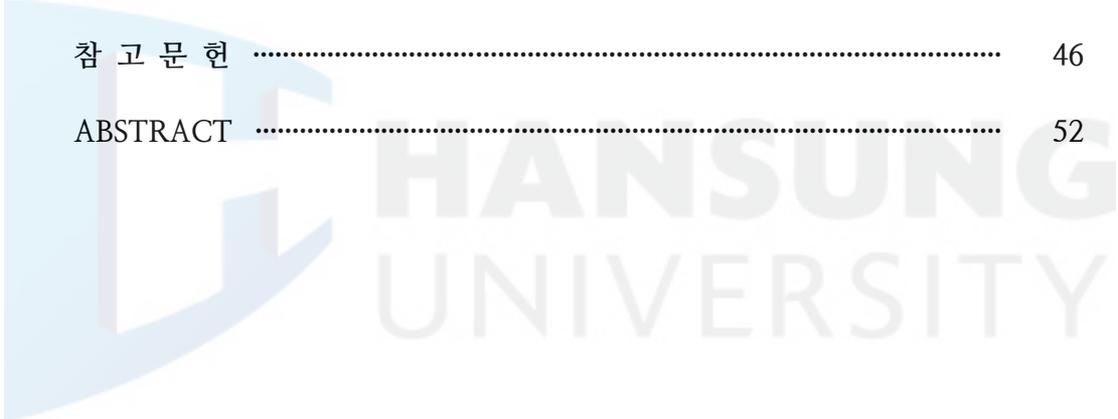


표 목 차

[표 2-1] Vision Transformer (ViT) 비교	8
[표 2-2] 객체 탐지 및 추출 방법 비교	11
[표 3-1] 연구 자료	14
[표 3-2] 영상 융합 레벨 비교	18
[표 4-1] Ground Truth 생성 자료	38
[표 4-2] KOMPSAT-3 광학 영상에 대한 계산 결과	42
[표 4-3] Wavelet 기반 융합 영상에 대한 계산 결과	42



그림 목 차

[그림 2-1] SAM 구조	8
[그림 2-2] Geo-SAM과 SAM의 구조 비교 및 처리 흐름도	10
[그림 2-3] ArcGIS를 활용한 SAM 기반 객체 분할 예시	12
[그림 3-1] 연구 지역	14
[그림 3-2] 연구 흐름도	16
[그림 3-3] 연구에서 사용한 Wavelet 기반 영상 융합 방법	20
[그림 3-4] Fine-tuning 개념 및 흐름도	21
[그림 3-5] Fine-tuning 방법	22
[그림 3-6] GLH-Water 데이터셋의 전 세계 위성 영상 분포도	23
[그림 3-7] GLH-Water가 제공하는 라벨 예시	24
[그림 3-8] 학습 데이터셋 생성 방법	26
[그림 3-9] GSV 이미지 전처리 과정	27
[그림 3-10] 데이터셋에 추가한 GSV 이미지	28
[그림 3-11] Fine-tuning 학습 손실 그래프	30
[그림 3-12] 합천호에 설치된 수상 태양광 패널	32
[그림 3-13] Geo-SAM 결과	33
[그림 3-14] Fine-tuning을 수행하지 않은 SAM의 결과	34
[그림 3-15] Fine-tuning 수행 후 SAM의 결과	35
[그림 4-1] 비교 자료	37
[그림 4-2] IoU와 mIoU의 개념 및 계산식	40
[그림 4-3] WIoU 개념 및 계산식	41

수 식 목 차

[수식 3-1] Wavelet 기반 융합 규칙	19
[수식 3-2] Wavelet 변환 공식	19
[수식 3-3] 역 Wavelet 변환 공식	19



제 1 장 서 론

제 1 절 연구 배경

최근 인공지능 기술의 발전으로 다양한 분야에서 대규모 사전 학습(pre-trained) 모델이 활용되고 있다. 특히 Meta AI가 2023년 4월에 발표한 Segment Anything Model (SAM)은 컴퓨터 비전 분야뿐만 아니라 의료, 산업, 원격탐사 등 다양한 영상 처리 분야에서 주목받고 있다(Kirillov et al., 2023). SAM은 추가 학습 없이 새로운 작업을 수행할 수 있는 제로샷(Zero-shot) 기반의 범용 이미지 분할 모델이다. 점(point), 경계 상자(bounding box), 마스크(mask), 문자(text) 등 다양한 유형의 프롬프트(prompt)를 입력하면 이미지 내 모든 객체를 자동으로 분할할 수 있으며 이는 기존 딥러닝 기반 분할 모델과 차별화되는 부분이다.

SAM은 1억 1,000만 장 이상의 영상과 10억 개 이상의 마스크로 구성된 Segment Anything 1 Billion (SA-1B) 데이터셋으로 학습되었으며, Vision Transformer (ViT)를 기반으로 한 생성형 구조로 다양한 크기 모델 아키텍처를 제공한다. ViT는 모델의 파라미터 수와 처리 능력에 따라 ViT-B (Base), ViT-L (Large), ViT-H (Huge)로 구분된다. 이미지 인코더(Image Encoder), 프롬프트 인코더(Prompt Encoder), 마스크 디코더(Mask Decoder) 세 가지 주요 구성요소로 이루어져 있으며, 다양한 객체와 상황에 대응할 수 있는 높은 범용성과 유연성을 확보하였다. 그러나 SAM은 일반적인 이미지로부터 학습되었기 때문에 고해상도 영상인 위성 영상에서 객체를 추출하는 데에는 분할 성능에 한계가 있다. 위성 영상은 다양한 센서 종류, 공간 해상도, 스펙트럼 특성에 따라 영상 표현 방법이 달라지며 특히 지형, 수체, 인공 구조물 등 다양한 객체가 포함된 복잡한 장면을 포함하고 있어 일반 이미지와는 다른 처리 방법이 요구된다.

위성 영상은 일반적으로 광학 영상과 Synthetic Aperture Radar (SAR) 영상으로 나뉜다. 이 두 영상은 지상을 관측하는 방법이 다르며, 각각 장단점

을 가지고 있다. 광학 영상은 색상 정보와 직관적인 해석이 가능하지만 구름이나 날씨, 환경 조건에 따라 정확도가 달라진다. 반면 SAR 영상은 기상이나 대기 조건에 관계없이 관측할 수 있으나, 노이즈로 인한 한계가 존재한다. 원격탐사 분야에서는 이러한 특성을 보완하기 위해 광학 영상과 SAR 영상을 융합하는 기법이 제안되고 있으며, 주파수 성분을 기반으로 융합 가능한 Wavelet 기반 이미지 융합 기법이 주목받고 있다.

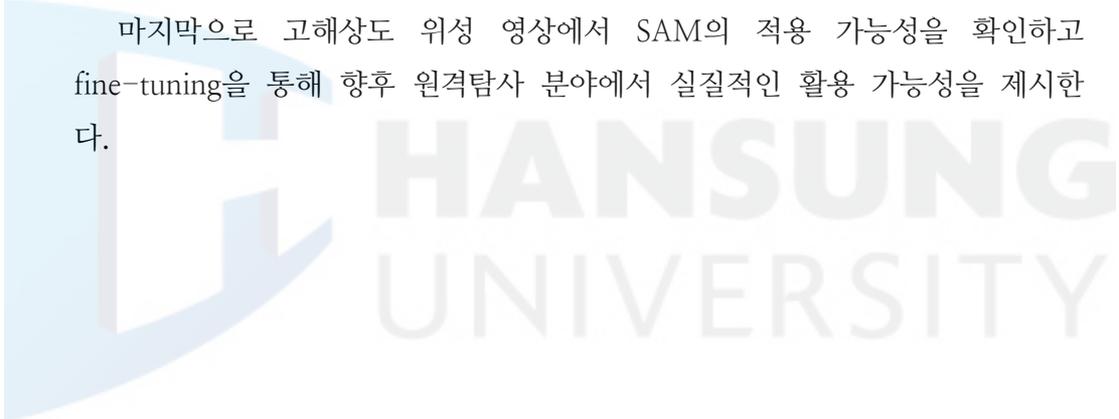
수체(waterbody)는 수자원 관리, 환경 변화 모니터링, 재해 예방 및 대응 등 다양한 분야에서 핵심이 되는 대상이다(Pietrucha-Urbanik, K. & Rak, J., 2023). 그러므로 수체를 탐지하는 데 위성 영상의 활용은 매우 유용하며, 이를 기반으로 한 활용 연구도 진행되고 있다. An and Rui (2022)는 고해상도 위성 영상을 활용해 수역 추출의 정밀도를 분석하고 개선 방안을 제시하였다. 또한 Xia et al. (2023)은 위성 영상을 이용하여 전 세계 수면 위 태양광 발전소의 분포를 파악하였으며 Hellweger et al. (2004)은 뉴욕 항만의 수질 조사를 위해 위성 영상을 적용한 사례를 소개하며 수체 모니터링에 실질적인 기여를 할 수 있음을 보여주었다. 이처럼 위성 영상을 활용한 수체 추출 연구도 활발히 진행되고 있다. 하지만 수체는 계절 및 강우량에 따라 불규칙한 형태를 가지며 저수지, 하천, 강, 바다 등 다양한 크기와 유형으로 존재한다. 이러한 특성은 수체 분할의 어려움을 가중시킨다. 따라서 최근 수체 추출을 위한 다양한 기계학습 및 딥러닝 기반 연구가 수행되고 있으며 QGIS 환경에서 지리공간 래스터 이미지와 SAM을 활용할 수 있는 플러그인인 Geo-SAM이 공개되었다(Zhao et al., 2023). 그러나 원격탐사 분야에서 SAM의 성능을 분석하고, 특정 성능을 향상시키기 위해 추가 학습인 미세조정(fine-tuning)을 적용하는 연구는 적은 실정이다.

제 2 절 연구 목적

연구의 목적은 고해상도 위성 영상을 활용하여 SAM(Segment Anything Model)으로 수체를 분할하고, fine-tuning을 통해 객체 분할 성능이 향상되는지를 확인하는 것이다. SAM은 일반 이미지 중심의 SA-1B 데이터셋으로

사전 학습되었기 때문에, 위성 영상에서도 객체 분할 기능이 효과적으로 작동하는지를 검토할 필요가 있다. 위성 영상은 광학 영상과 SAR 영상으로 나뉘며, 각각 색상 정보와 기상 영향을 받지 않는 관측 정보를 제공하는 상호 보완적인 특성을 가진다. 이에 KOMPSAT-3 광학 영상과 KOMPSAT-5 SAR 영상을 융합하여 실험을 진행하고, 단독 영상과 비교하여 어떤 영상이 더 정밀한 분할 결과를 제공하는지 분석하였다. 사전 학습된 SAM은 다양한 객체에 대응할 수 있지만, 위성 영상 기반의 학습 경험이 부족하고, 불규칙한 자연 경계에 대한 인식 성능이 떨어질 수 있다. 따라서 수채 중심의 학습 데이터셋을 구성하고, 프롬프트 인코더와 마스크 디코더를 중심으로 fine-tuning을 수행하였다. 이후 수채 분할 결과에 대해 위치 정확도와 면적 기반의 일치도를 평가하기 위해 다양한 비교 자료를 활용하여 성능을 검증하였다.

마지막으로 고해상도 위성 영상에서 SAM의 적용 가능성을 확인하고 fine-tuning을 통해 향후 원격탐사 분야에서 실질적인 활용 가능성을 제시한다.



제 2 장 객체 분할 기술 및 연구 동향

제 1 절 객체 탐지 및 분할 기술

객체 탐지(Object Detection)와 분할(Segmentation)은 컴퓨터 비전 분야에서 핵심적인 과제로, 다양한 이미지나 비디오에서 객체를 식별하고 위치를 파악할 수 있도록 한다(Wang X, 2016). 최근 딥러닝(Deep Learning) 기술의 발전으로 이러한 작업의 정확도와 효율성이 크게 향상되었으며, 더 나은 성능을 위해 탐지와 분할을 통합하는 방법 등 다양한 방법이 늘어나고 있다(Girshick et al., 2015).

객체 탐지는 일반적으로 경계 상자를 활용하여 객체가 어디에 존재하는지 식별하고, 해당 위치에 있는 객체가 정의된 레이블 중 어떤 레이블에 속하는지 파악한다. 이러한 기술은 자율주행 차량의 차량 감지 및 추적, CCTV 기반 이상행동 추적, 로봇틱스 분야, 불량 검출 등 다양한 분야에서 실용적으로 활용된다. 딥러닝 기반 객체 탐지 알고리즘은 구조에 따라 Classification-based 방법과 Regression-based 방법으로 나뉜다. Classification-based 알고리즘은 먼저 Region Proposal 단계에서 Region of Interest (ROI)를 생성 후, 각 영역에 대하여 Convolution Neural Network (CNN)를 통해 객체의 유무를 분류하는 방법이다. 이 방법은 비교적 정확한 탐지가 가능하지만, 각 ROI마다 별도의 순방향 전파(forward pass)를 수행해야 하므로 계산량이 크고 속도가 느린 단점이 있다. 대표적인 모델로는 R-CNN(Girshick et al., 2014)이 있으며 이후 R-CNN(Ren et al., 2015), Faster R-CNN으로 개선되었다. Regression-based 알고리즘은 ROI 추출 단계를 생략하고, 한 번의 연산으로 전체 이미지에서 객체 클래스와 경계 상자를 동시에 예측한다. 때문에 계산 속도가 빠르고, 실시간 처리에 적합하다. 대표적인 예로 You Only Look Once (YOLO)는 이미지를 한 번에 처리하여 빠른 예측을 가능하게 하고, Single Shot MultiBox Detector (SSD)는 다양한 크기의 디폴트 박스(default box)를 활용하여 객체를 감지하도록 설계되었다.

분할은 이미지를 픽셀 단위로 구분하는 작업으로 컴퓨터 비전의 한 분야이다. 이는 크게 의미 분할(Semantic Segmentation)과 인스턴스 분할(Instance Segmentation)로 나뉜다. 의미 분할이란 이미지에 포함된 모든 픽셀에 대해 예측하는 것이다. 또한 분할된 이미지를 구분할 수 있으나 영역이 겹치는 경우 같은 영역으로 인식하여 각 객체를 구분할 수는 없다. 예로 Fully Convolutional Network (FCN), 대칭 인코더-디코더 구조를 활용한 U-Net, 복잡한 배경에서도 공간적 세부 정보를 보존하는 DeepLabV3+ 등이 있다. 반면 인스턴스 분할은 객체를 구분하는 작업이 객체 탐지와 비슷한 수준이며, 같은 영역의 객체가 구분 가능하다는 점에서 의미 분할과 차이점이 있다. 대표적인 예로, Mask R-CNN는 특정 위치를 예측하고 동시에 정확한 마스크를 생성할 수 있어 정밀한 분할에 적합하다.

이러한 객체 탐지 및 분할 기술은 의료 영상, 자율주행 차량, 고해상도 위성 영상 분석 등 다양한 분야에 활용되고 있으며, 최근에는 사전 학습 범용 모델을 사용한 연구가 진행되고 있다(Zhou et al., 2024).

제 2 절 Segment Anything Model (SAM)

SAM은 2023년 Meta AI에서 발표한 범용 객체 분할 모델로, 사전 학습 없이 다양한 이미지에 대해 분할을 수행할 수 있는 제로샷 기능을 제공한다. 기존의 클래스 기반 객체 분할 방식과 달리, SAM은 사용자가 입력한 프롬프트를 사용하여 객체를 분할하는 Promptable Segmentation 방식을 사용한다. 프롬프트는 점, 경계 상자, 문자 등 다양한 형태로 구성된다. SAM은 학습 단계에서 다양한 형태의 모호한 프롬프트(ambiguous prompt)를 함께 학습했기 때문에 명확하지 않은 위치나 객체에도 유효한 마스크를 생성할 수 있다.

SAM은 [그림 2-1]과 같이 이미지 인코더, 프롬프트 인코더 그리고 마스크 디코더 세 가지 모듈로 구성된다. 먼저 이미지 인코더는 Vision Transformer (ViT)를 기반으로 사전 학습된 Masked Autoencoder (MAE)를

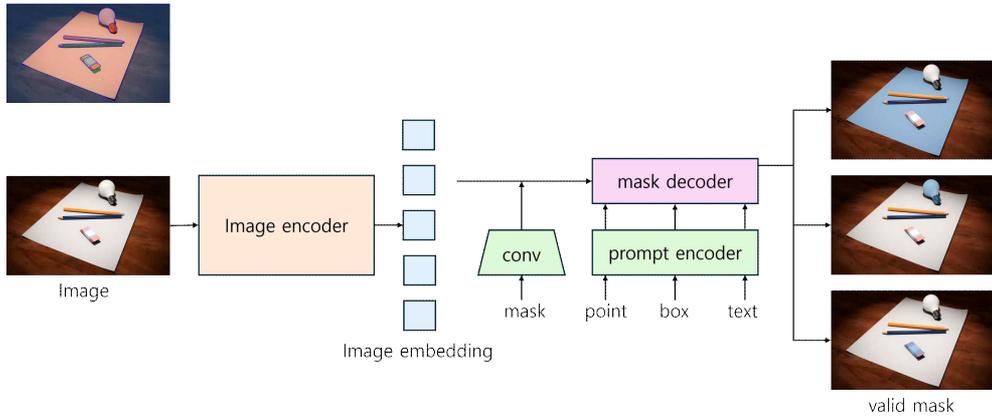
통해 이미지로부터 시각적 특징을 추출한다. ViT는 트랜스포머(transformer)의 깊이, 숨겨진 차원 수(hidden dimension), 그리고 전체 파라미터 수에 따라 ViT-B (Base), ViT-L (Large), ViT-H (Huge) 순으로 구분된다 (Dosovitskiy et al., 2020). [표 2-1]은 이를 요약하여 정리한 표이다. ViT는 이미지를 일정 크기의 패치로 분할한 뒤 트랜스포머 구조에 입력하는 방법으로 작동하기 때문에 전체 이미지를 파악할 수 있으며 객체 사이의 관계와 구조를 효과적으로 학습할 수 있다. 프롬프트 인코더는 사용자가 입력한 프롬프트를 sparse 또는 dense 방법을 사용하여 임베딩을 추출한다. Sparse 방법에는 점, 경계 상자, 문자 세 가지로 구성되며, 점과 경계 상자는 입력 이미지 속 특정한 좌표를 기준으로 positional encoding을 적용하여 임베딩 된다. 문자는 Contrastive Language-Image Pretraining (CLIP) 모델의 텍스트 인코더를 사용하여 의미를 벡터화한다. Dense 방법에는 마스크 한 가지로 구성되며 이는 합성곱(convolution) 연산을 통해 spatial feature로 변환되고 이미지 임베딩과 정밀한 결합이 가능하다. 마지막으로 마스크 디코더는 이미지 임베딩과 프롬프트 임베딩을 결합하여 분할 마스크를 생성한다. 트랜스포머 기반 구조로 양방향 self-attention과 cross-attention 메커니즘을 통해 이미지와 프롬프트 간 정보를 상호 교환하며 정교한 결과를 생성한다. 또한 단일 마스크 생성 방법이 아닌, Dynamic Mask Prediction 방법으로 하나의 프롬프트에 대해 여러 개의 마스크를 동시에 생성할 수 있다. 이는 앞서 설명했듯 SAM이 모호한 프롬프트에 대해 학습되었으며 대응할 수 있기 때문이다. 이렇게 예측된 마스크는 probability map 또는 이진 마스크(binary mask) 형태로 나타난다.

최근 객체 탐지와 객체 추출을 목적으로 SAM을 활용하는 연구가 활발히 진행되고 있다. Carraro et al. (2023)은 농업 분야에서 SAM을 활용하여 분할된 작물 마스크를 분석하였고, Mazurowski et al. (2023)은 의료 영상 분석에서 SAM의 대화형 프롬프트 처리 성능을 평가하여 박스형 프롬프트의 우수성을 발표하였다. Yang et al. (2024)과 Giannakis et al. (2024)은 각각 태양광 패널 탐지와 지구과학 자료 기반 분화구 분석에 SAM을 적용하여 실험한 결과를 발표했다. Osco et al. (2023)은 대용량 위성 영상에서 자동화된 프롬프트 처리 가능성을 분석하며 SAM의 가능

성과 한계를 지적하였고, Wu and Osco (2023)는 공간 데이터 처리를 위한 파이썬 패키지를 발표하여 SAM의 활용성을 높였다. Yan et al. (2023)은 광학 영상과 SAR 영상을 대상으로 SAM과 기존 딥러닝 기법의 성능을 비교하였다.

실무 응용 측면에서도 Zhang et al. (2023)은 Sentinel-2 영상을 활용해 SAM 기반의 경작지 분류 및 매핑 기능을 통합하는 농업 모니터링 시스템을 개발하였으며, Ren et al. (2023)은 양식업 분야에서 불법 조업 감시를 위한 위성 영상 기반 SAM의 적용 사례를 보고했다. 그러나 Moghimi et al. (2024)은 강 유역 탐지 실험에서 SAM의 정확도가 딥러닝 알고리즘보다 낮고 일관성을 보장하지 못한다고 비판하였다. Gui et al. (2024)은 SAM의 제로샷 학습 능력이 향후 지구 관측 및 객체 탐지 기술의 핵심이 될 것이라 평가하였으나, Ren et al. (2024)은 실사용 환경에서 일부 객체의 분할 성능 저하를 지적하며 fine-tuning의 필요성을 강조했다. Xue et al. (2024)은 ViT 모델의 fine-tuning 처리 결과가 SAM의 정확도에 영향을 미친다는 점을 발표하였다.

SAM의 범용성과 효율성이라는 강점에도 불구하고 일부 객체의 특성 때문에 fine-tuning 없이는 정확한 추출이 어려울 수 있다. 따라서 이러한 한계를 극복하고 활용도를 높이기 위해 향후 정교화된 실험 및 연구가 필요하다.



[그림 2-1] SAM의 구조

[표 2-1] Vision Transformer (ViT) 비교

모델	계층 수 (Layers)	숨겨진 차원의 수 (Hidden size D)	파라미터 (params)	특징
ViT-B	12	768	86M	빠른 처리 속도, 중간 성능
ViT-L	24	1024	307M	균형잡힌 성능
ViT-H	32	1280	632M	높은 성능, 높은 연산 자원 필요

제 3 절 Geo-SAM

Geo-SAM은 Zhao et al.,(2023)이 발표한 오픈소스 QGIS 플러그인으로서, SAM을 기반으로 대규모 지리공간 이미지를 분할하고 라벨링할 수 있도록 지원한다. Geo-SAM은 SAM의 기본 구조를 QGIS 환경에 맞게 최적화하여 구현한 것으로, [그림 2-2]는 SAM과 Geo-SAM의 처리 구조 및 흐름을 비교하여 시각적으로 나타낸 것이다.

Geo-SAM은 크게 이미지 인코딩 부분과 대화형 분할(Interactive Segmentation)로 구성된다. 먼저 이미지 인코딩 부분은 SAM의 이미지 인코더를 사용하여 영상의 특징을 생성하고 저장한다. 이러한 과정은 이미지당 한

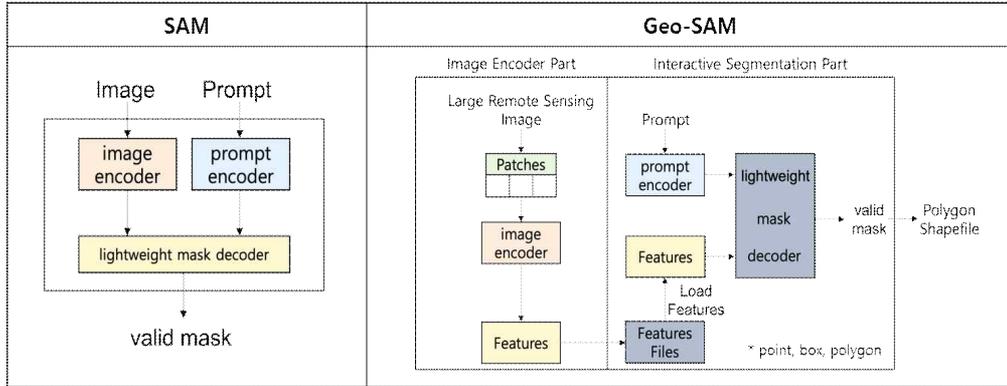
번만 수행되면 된다. 이후 생성된 특징은 대화형 분할 기능을 통해 사용자의 입력, 즉 프롬프트를 통해 실시간으로 분할이 가능하다. 프롬프트는 Foreground point (FG), Background point (BG), 경계 상자(BBox) 세 가지 유형이 제공된다. FG는 원하는 영역을 나타내는 점이고, BG는 원하지 않는 영역을 나타내는 점이다. 마지막으로 BBox는 객체의 경계를 제한하는 경계 상자이다. 위성 영상의 경우 다양한 색상을 가지고 있기 때문에 점이나 경계 상자를 구분하기 어려울 수 있다. 따라서 색상은 사용자가 직접 조정할 수 있으며 스타일 탭에서 변경 가능하다.

Geo-SAM은 직관적인 그래픽 사용자 인터페이스(GUI: Graphical User Interface)를 제공하여 QGIS 환경에서 쉽게 사용할 수 있도록 설계되었다. 또한 프롬프트 기반의 실시간 분할 기능을 통해 작업의 효율성과 정확도를 높일 수 있다. 하지만 SAM의 구조상 한 번에 하나의 객체만 분할할 수 있기 때문에, 결과를 별도로 저장해야 하는 반복적인 작업이 요구된다.

기본적으로 SAM은 R, G, B 세 가지 밴드의 영상을 입력하도록 설계되었지만, Geo-SAM은 SAR 영상, Grayscale 영상, NDVI, NDWI 등 1~2 밴드로 구성된 다양한 영상에도 적용할 수 있다.

Geo-SAM은 GitHub 저장소에서 zip 파일을 수동으로 다운로드하거나, git clone 명령어를 통해 저장소를 복제하는 두 가지 방법으로 설치할 수 있다. 또한 Geo-SAM을 사용하려면 SAM 모델의 체크포인트인 ViT가 미리 설치되어 있어야 하며, ViT-L 모델을 백본(backbone)으로 사용한다.

QGIS의 공간 분석 기능과 SAM의 객체 분할 기능을 결합한 Geo-SAM은 대규모 지리공간 데이터를 보다 효율적으로 처리할 수 있도록 돕는다. 현재도 지속적인 기능 개선과 안정성 확보를 위한 업데이트가 진행 중이며, 원격탐사 및 공간 정보 분석 분야에서 실용화 가능성이 확대되고 있다.



[그림 2-2] Geo-SAM과 SAM의 구조 비교 및 처리 흐름도

제 4 절 원격탐사 기반 객체 추출 연구

최근 원격탐사(Remote Sensing) 분야에서는 고해상도 위성 영상 확보와 함께 객체 탐지 기술이 발전하면서 건물, 도로 등 다양한 지상 객체를 추출하는 연구가 활발해지고 있다. 위성 영상은 넓은 지역을 한 번에 관측할 수 있고 높은 공간 및 스펙트럼 해상도를 제공하여 다양한 분야에서 중요한 자료로 활용된다.

기존 연구의 대부분은 광학 영상과 SAR 영상을 각각 활용하였으나, 두 영상을 결합하여 분할 성능을 향상시키고 복잡한 영상에서도 사용하려는 시도가 증가하고 있다. Kong et al.(2021)은 Dense-UGAN과 Gram-Schmidt 변환을 통해 광학 영상의 스펙트럼 정보와 SAR 영상의 텍스처 정보를 융합하는 방법을 제안하였다. 또한 Liu et al.(2023)은 딥러닝 기반 융합 기법을 사용하여 의미 분할의 정확도를 향상시키고 단일 센서의 한계를 보완하였다.

위성 영상의 종류와 사용 목적에 따라 객체 탐지 및 추출에 적용할 수 있는 기법은 다양하며, 일반적으로 전통적인 알고리즘 기반 방법과 딥러닝 기반 방법으로 나뉜다. 알고리즘 기반 방법은 대부분 스펙트럼의 임계값을 설정하거나 통계적 특성을 분석하는 방식으로 수행된다. 반면 딥러닝 기반 방법은 객체의 형태와 경계 정보를 계층적으로 학습할 수 있는 구조를 가진다. [표 2-2]는 객체 탐지 및 추출 방법을 비교하여 작성한 표이다.

최근에는 두 방법을 결합하는 하이브리드 방법(Hybrid Method) 또는 SAM처럼 거대한 데이터로 사전 학습된 모델을 활용해 다양한 객체를 빠르고 자동으로 분할하는 기술이 주목받고 있다. 하지만 복잡한 형태나 작은 객체에 대한 처리에는 한계가 있어, fine-tuning으로 성능을 개선하는 연구도 진행 중이다. [표 2-2]는 이러한 세 가지 방법의 특징을 비교한 표이다.

또한 Geo-SAM 플러그인 외에도 SAM을 활용한 플랫폼 기반 연구와 서비스가 확대되고 있다. ArcGIS는 건물 윤곽 추출, 북극 해양생물 탐지, 수체 추출 등 다양한 원격탐사 응용 분야에서 SAM을 이용해 고정밀 객체 분할을 제공하며, 이를 통해 대규모 영상 데이터 분석에 SAM 활용성을 높이고 있다.

[표 2-2] 객체 탐지 및 추출 방법 비교

비교 항목	알고리즘 기반 방법	딥러닝 기반 방법	SAM (Segment Anything Model)
대표 기법	NDWI, NDVI, SVM, Random Forest	U-Net, DeepLab, Mask R-CNN	Prompt 사용 Point, Box, Text
학습 필요성	학습 불필요, 경우에 따라 소규모 데이터 학습 필요	대규모 학습 데이터 필요	추가 학습 없이 제로샷 기능 사용
처리 속도	빠른 처리 속도, 경우에 따라 느려지기도 함	느린 처리 속도, 연산 자원 요구	빠른 처리 속도, 실시간 적용 가능
데이터 의존성	스펙트럼 밴드에 따라 성능 결정, 환경 변화에 민감	학습 데이터 품질과 양에 의존, 과적합 가능성 있음	다양한 데이터 사용 가능
한계점	복잡한 환경에서 정확도 저하, 다목적성 부족	대규모 데이터와 긴 학습 시간 요구, 과적합 가능성 존재	작은 객체 및 복잡한 구조에서 성능 저하

ArcGIS pretrained models

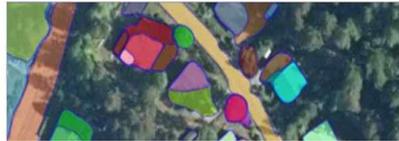
Search ArcGIS pretrained models

Get started Imagery Point cloud Text

Imagery / Segment Anything Model (SAM)

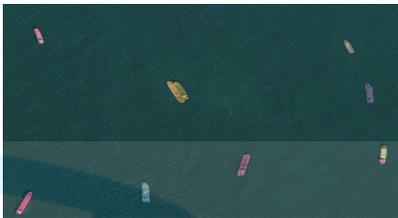
- > Arctic Seal Detection
- > Building Footprint Extraction - Africa
- > Building Footprint Extraction - Australia
- > Building Footprint Extraction - China

Introduction to the model



In this topic

- License requirements
- Model details
- Access and download the model
- Release notes



[그림 2-3] ArcGIS를 활용한 SAM 기반 객체 분할 예시

제 3 장 연구 자료 및 방법

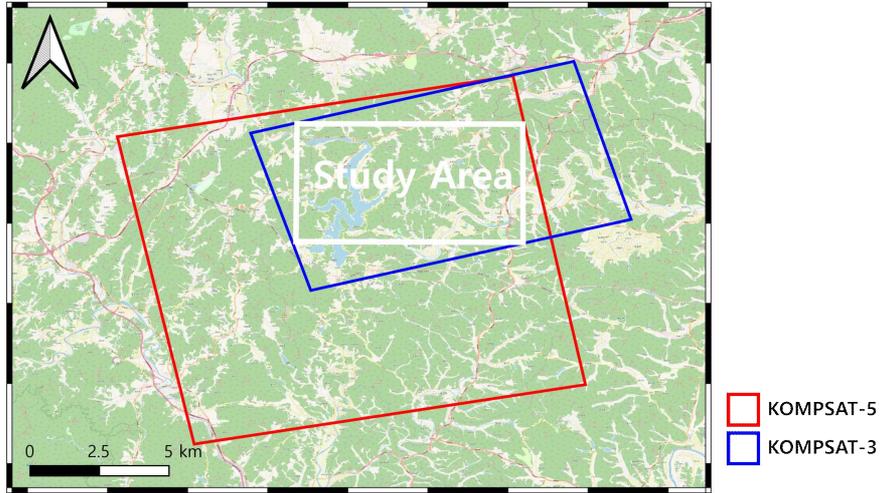
제 1 절 연구 지역 및 연구 자료

연구 지역은 경상남도 합천군 합천호 일대로, 다양한 크기의 수체가 분포된 지역이다[그림 3-1]. 특히 이 지역은 낙동강의 지류인 황강과 합천호를 중심으로 하며, 장계 저수지와 정양 저수지가 존재한다. 합천호는 황강을 막아 형성된 인공호수로, 면적은 약 25.95km²이며 만수위는 176m, 저수량은 약 7억 9000만 톤으로 홍수 조절 및 농업용수 공급 등 중요한 역할을 한다. 황강은 길이 111.00km의 낙동강 지류로, 합천군을 중심으로 흐르고 있으며 주변의 산악 지형과 연결되어 계절 및 기상 조건에 따라 수위의 변화가 크다. 또한 장계 저수지와 정양 저수지는 비교적 작은 크기의 수체로 위성 영상에서 경계가 명확하지 않은 특성을 보인다. 이처럼 다양한 크기와 형태의 수체가 분포하는 연구 지역은 객체 탐지 및 추출 연구를 위한 다양한 실험 조건을 제공한다.

연구에 사용된 자료들은 국내 고해상도 광학 영상인 KOMPSAT-3와 SAR 영상인 KOMPSAT-5 영상으로 각각 2023년 3월 10일과 2023년 3월 9일에 취득된 GeoTIFF 형식의 영상들이다. 이들의 영상 촬영 일이 하루 차이로 매우 유사한 시기에 촬영되었으며 비교 및 융합하기에 유리하다. 또한 QGIS 환경에서 Geo-SAM 플러그인과 점 프롬프트를 사용하여 수체 Shapefile을 추출하였다. 추가적으로 QuickOSM 플러그인을 이용하여 수체에 대한 Shapefile 데이터를 추출하였으며, 국토리지정보원(NGII: National Geographic Information Institute)에서 제공하는 축척 1/5,000의 면(polygon) 형태의 수계 및 하천 경계(N3A_E0010001)와 호수/저수지 경계(N3A_E0052114) Shapefile 자료를 활용하였다. 이러한 자료들은 KOMPSAT-3와 KOMPSAT-5 영상의 수체 추출 결과와 비교 분석 및 평가를 위한 비교 자료로 사용되었다. [표 3-1]은 연구에 사용한 자료들을 요약한 표이다.

127.741452, 35.751404

128.342475, 35.751404



127.741452, 35.331512

128.342475, 35.331512

[그림 3-1] 연구 지역

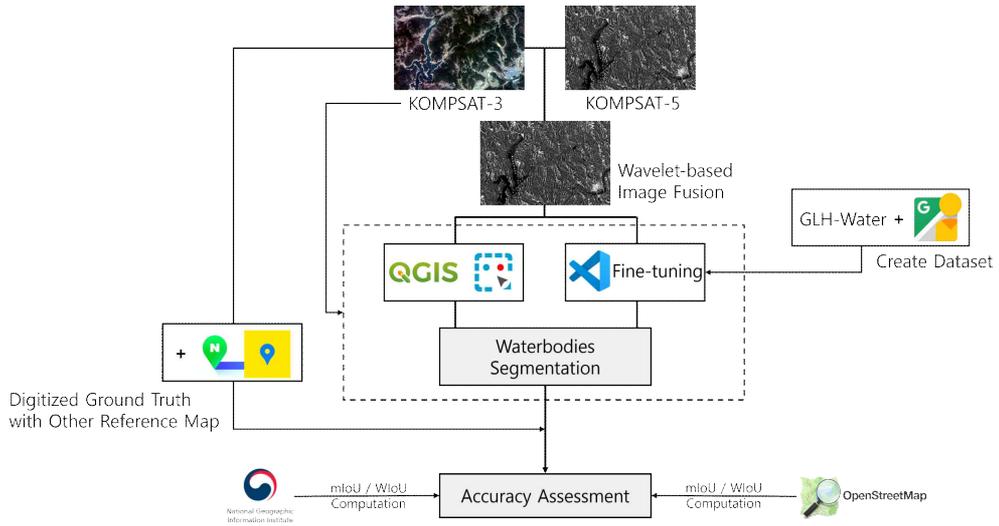
[표 3-1] 연구 자료

자료		파일 종류	정보	취득 날짜
KOMPSAT-3		GeoTIFF	- 제품 유형: R, G, B - 제품 수준: level1G - 해상도 : 2.8m	2023.03.10
KOMPSAT-5			- 제품 유형: HH - 제품 수준 : level1A - 해상도 : 1m	2023.03.09
Geo-SAM		Shapefile	- QGIS 플러그인 - Version : 1.3.1 - SAM 기반 지리공간 객체 분할	2024.12.24
QuickOSM			- QGIS 플러그인 - Version : 2.2.3 - Key/Value: Natural/Water	2024.05.22
NGII	N3A_E0010001		- 축척 : 1/5,000 - 형식 : Polygon - 분류 : 수계-하천 경계	2024.02.13
	N3A_E0052114	- 축척 : 1/5,000 - 형식 : Polygon - 분류 : 수계-호수/저수지	2024.02.13	

제 2 절 연구 방법

고해상도 위성 영상을 활용한 수체 추출을 위해 [그림 3-2]는 연구의 자료 처리 방법과 연구 흐름도를 정리한 것이다. 우선, 광학 영상인 KOMPSAT-3와 SAR 영상인 KOMPSAT-5에 대해 영상 전처리가 필요하다. 특히 SAR 영상은 노이즈 및 스펙클 현상으로 인해 원본 영상을 그대로 사용할 수 없다. 이러한 문제점을 해결하기 위해 European Space Agency (ESA)에서 무료로 배포하는 Sentinel Application Platform (SNAP)을 활용하여 SAR 영상에 대한 전처리를 수행하였다.

전처리된 KOMPSAT-3와 KOMPSAT-5 영상을 대상으로 Wavelet 기반 이미지 융합 기법을 사용하여 이미지 융합을 진행하였다. 이후 KOMPSAT-3 영상과 융합 영상에 QGIS 플러그인인 Geo-SAM을 사용하여 수체 추출을 진행하였으며, 이때 ViT-L 모델과 세밀한 객체 추출을 위해 점 프롬프트를 활용하였다. 또한 수체 객체 특성에 적합하도록 fine-tuning을 통해 모델 성능을 개선하여 SAM을 수행하였다. 이 과정에서 위성 영상과 정밀한 주석 데이터로 구성된 GLH-Water 데이터셋을 활용하였다. GLH-Water 데이터셋을 단독으로 사용하지 않고, GLH-Water 데이터셋에 Google Street View (GSV)에서 수체를 분할한 이미지를 추가하여 새로운 데이터셋을 생성하였다. 추출 결과는 QuickOSM 플러그인으로부터 얻은 수체 객체 자료와 NGII의 수계 수치 자료 그리고 KOMPSAT-3 영상과 네이버 지도, 카카오맵, OSM을 참조하여 생성한 Ground Truth 자료와 비교하여 정확도 평가를 진행하였다. 정확도 평가는 추출된 객체와 실제 객체 간의 교차 면적과 전체 면적을 비교하여 정확도를 평가하는 지표인 Mean Intersection over Union (mIoU) 과 객체의 크기를 고려하여 큰 객체에 더 큰 가중치를 부여하는 Weighted IoU (WIoU)를 사용하였다.



[그림 3-2] 연구 흐름도

제 3 절 이미지 융합 기법

이미지 융합은 두 개 이상의 영상을 결합하는 과정으로, 정보성이 뛰어나 원격탐사 분야에서 중요한 역할을 한다(Byun et al., 2013). 다양한 원격탐사 위성의 발전으로 여러 위성이 수집한 영상에 대해 상호 보완적인 정보를 얻을 수 있게 되었으며 이미지 융합 기법의 필요성이 더욱 커지고 있다(Kulkarni and Rege, 2020). 특히 광학 영상과 SAR 영상의 장점을 결합하여 영상을 융합하면 객체 탐지 및 분할 성능을 향상시킬 수 있다. 광학 영상은 높은 공간 해상도와 풍부한 색상 정보를 제공하지만, 기상 조건, 대기 상태, 그림자 등 외부 환경의 영향을 받기 쉽다. 반면 SAR 영상은 기상 조건이나 촬영 시간에 구애받지 않고 영상을 획득할 수 있다는 장점이 있으나, 노이즈와 스펙클 현상으로 경계 인식이 어렵다는 한계가 있다.

1) 원격탐사 기반 이미지 융합 기법

광학 영상과 SAR 영상을 융합하는 방법은 다양하며 정보 추상화 수준에 따라 세 가지 레벨로 분류할 수 있다(Rane et al., 2017). 픽셀 레벨 융합, 특

징 레벨 융합, 그리고 결정 레벨 융합이다. [표 3-2]는 세 가지 영상 융합 레벨을 비교한 표이다.

이미지 융합에는 다양한 기술들이 사용되며, 대표적으로 라플라스 변환 기반 융합, Wavelet 기반 융합, 주성분 분석(PCA: Principal Component Analysis) 등이 있다. 최근 딥러닝 기술의 발전으로 CNN 기반 융합 연구도 활발해지고 있으며 또한 Intensity-Hue-Saturation (IHS) 기반 융합, 팬샤프닝(pansharpening) 등 여전히 고해상도 영상 처리에서 필수적인 역할을 한다.

Wavelet 기반 이미지 융합은 다중 센서에서 획득한 영상을 다중 해상도 분석을 통해 결합하는 기법으로, 공간 및 주파수 정보를 동시에 고려할 수 있다는 장점이 있다(Li et al., 1995). 영상의 경계와 세부 구조를 유지하면서도 노이즈를 최소화하고 융합을 수행하기 때문에 원격탐사 분야뿐만 아니라 공간 정보 처리에도 활용되고 있다.

PCA 기반 융합은 불필요한 정보와 노이즈를 제거하면서 필요한 특징은 유지하여 활용하는 기법이다. Chamundeeswari et al. (2009)은 PCA의 장점으로 정보의 단순화와 노이즈 제거를 언급하며, 광학 영상과 SAR 영상을 효과적으로 결합할 수 있다고 제시하였다.

최근 CNN 기반 융합은 원격탐사 영상 처리 분야에서 자주 사용되고 있으며, 특히 SAR과 Optical 데이터의 융합 및 분류에 효과적으로 적용되고 있다(Shakya et al., 2020). CNN은 Deep Neural Network (DNN)의 한 종류로 여러 개의 합성곱 층(convolution layer), 풀링 층(pooling layer), 완전 연결 층(fully connected layer)로 구성된다. 이는 학습에 적합한 구조이며 뛰어난 이미지 인식 성능을 보인다.

IHS 기반 융합은 다중 스펙트럼 영상의 색상 정보를 유지하면서 공간 해상도를 향상시키기 위해 자주 사용되는 기법이다. RGB 이미지를 IHS 색 공간으로 변환한 후, 명도 채널을 고해상도 영상으로 대체하고 다시 RGB 공간으로 복원하는 방법이다. 또한 팬샤프닝은 고해상도 팬크로매틱(panchromatic) 영상과 저해상도 다중 스펙트럼 영상을 결합하여 고해상도 다중 밴드 영상을 생성하는 기법으로, 팬크로매틱 영상 수준의 공간 해상도를 유지하면서 다중 스펙트럼 영상의 색상 정보를 스펙트럼 정보로 보존한다. 이

기법은 고해상도 이미지를 신속하게 얻을 수 있어 원격탐사 및 지리정보 분석에서 널리 활용되지만, 스펙트럼 왜곡이나 세부 정보 손실의 한계가 존재한다(Yehia et al., 2019).

[표 3-2] 영상 융합 레벨 비교

융합 레벨	특징	추가 정보
픽셀 레벨 융합	<ul style="list-style-type: none"> - 가장 낮은 수준의 영상 융합 - 다른 레벨의 영상 융합의 기반 - 다른 레벨에는 없는 상세한 정보를 포함 	데이터 융합 정확도가 가장 높음
특징 레벨 융합	<ul style="list-style-type: none"> - 중간 수준의 이미지 융합 - 중요한 정보를 유지하면서 정보를 압축 	<ul style="list-style-type: none"> - 실시간 처리에 유리 - 다양한 센서로부터 원시 정보를 추출한 뒤 분석 및 처리
결정 레벨 융합	<ul style="list-style-type: none"> - 가장 높은 수준의 이미지 융합 - 가장 복잡하며 이미지 등록의 필요성이 낮음 	명령 및 제어를 위한 의사 결정 근거 제공

2) 연구에서 사용한 Wavelet 기반 융합

KOMPSAT-3 광학 영상과 KOMPSAT-5 SAR 영상을 융합하여 보다 정밀한 수체 추출을 위해 Wavelet 기반 영상 융합 기법을 선택하였다. Wavelet 기반 융합은 입력 영상을 Wavelet 변환을 통해 다중 해상도 계수로 분해한 뒤, 각 계수 대역(저주파와 고주파)에서 융합 규칙을 적용하여 결합하고, 역 Wavelet 변환(Inverse Wavelet transform)을 통해 융합 영상을 재구성하는 절차를 포함한다. 이 과정은 공간 및 주파수 정보를 모두 반영할 수 있어, 광학 영상의 색상 및 구조 정보와 SAR 영상의 경계 및 질감을 동시에 유지할 수 있다는 장점을 갖는다.

[수식 3-1]의 융합 규칙을 기반으로 Wavelet 이미지 융합을 진행하였으며, W_{MS}^{LOW} 는 광학 영상의 저주파 계수 W_{SAR}^{high} 는 SAR 영상의 고주파 계수를 나타낸다. 광학 영상의 색상 및 명암 정보는 저주파(LL) 성분, 질감과 경계 표현은 SAR 영상의 고주파(HL, LH, HH) 성분 각각 매칭되도록 설정하였다. 또한 영상 융합 과정에서 영상 간 위치 정합도와 해상도 보정은 필수적이므로 ESA의 SNAP을 활용하여 전처리를 수행하였다.

$$W_{fused}(MS, SAR) = \frac{W_{MS}^{LOW}(MS, SAR) \text{for low-freq components}}{W_{SAR}^{high}(MS, SAR) \text{for high-freq components}}$$

[수식 3-1] Wavelet 기반 융합 규칙

연구에서 사용한 Wavelet 기반 융합 기법의 방법 및 흐름도는 [그림 3-3]에 나타내었다. 먼저 입력 영상을 모두 다중 해상도로 분해한다. 이후 융합 규칙에 따라 저주파와 고주파 계수를 선택하고 목적에 맞게 조합한다. 마지막으로 역 Wavelet 변환을 통해 최종 융합 영상을 생성한다. 이 기법은 영상의 주파수 정보를 고려하여 영상의 세부 구조를 유지하면서도 노이즈를 최소화할 수 있다는 장점이 있으며, 서로 다른 센서 특성을 반영할 수 있는 원격탐사 분야에서 높은 활용 가능성을 갖는다.

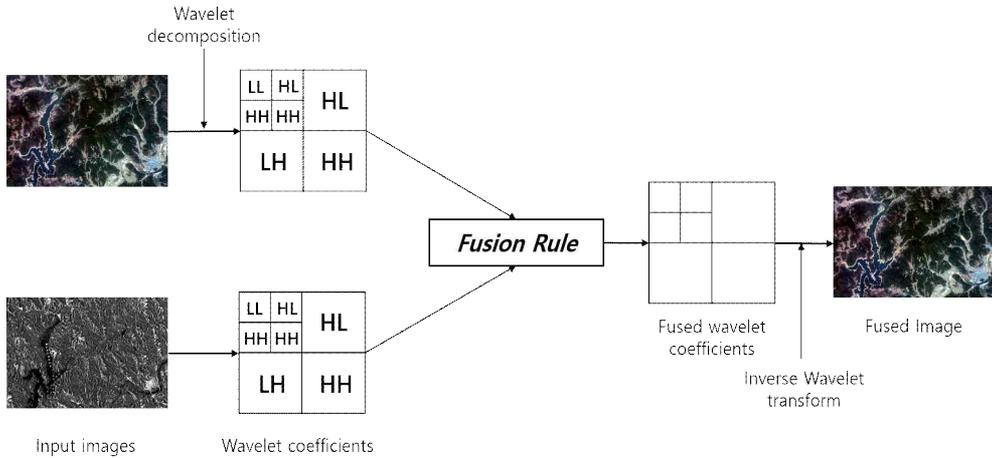
$$W_{ms} = \text{Wavelet Transform}(I_{ms})$$

$$W_{SAR} = \text{Wavelet Transform}(I_{SAR})$$

[수식 3-2] Wavelet 변환 공식

$$I_{fused} = \text{InverseWaveletTransform}(W_{fused})$$

[수식 3-3] 역 Wavelet 변환 공식



[그림 3-3] 연구에서 사용한 Wavelet 기반 영상 융합 방법

제 4 절 Fine-tuning 및 데이터셋

1) Fine-tuning

Fine-tuning은 사전 학습된 모델을 기반으로 새로운 데이터셋을 사용하여 모델의 일부 또는 전체 가중치를 미세하게 조정하여 특정 도메인이나 작업에 적합하도록 최적화하는 과정이다(Cheng et al., 2023). 이는 새로운 모델을 구축하거나 기존 모델을 처음부터 학습시키는 것에 비해 학습 시간을 줄이면서도 높은 성능을 확보할 수 있다(Feng et al., 2023). SAM과 같은 대규모 비전 모델은 사전 학습을 통해 범용 분할 성능은 확보하였으나, 특정 도메인에 최적화된 성능을 위해서는 fine-tuning이 필수적이다.

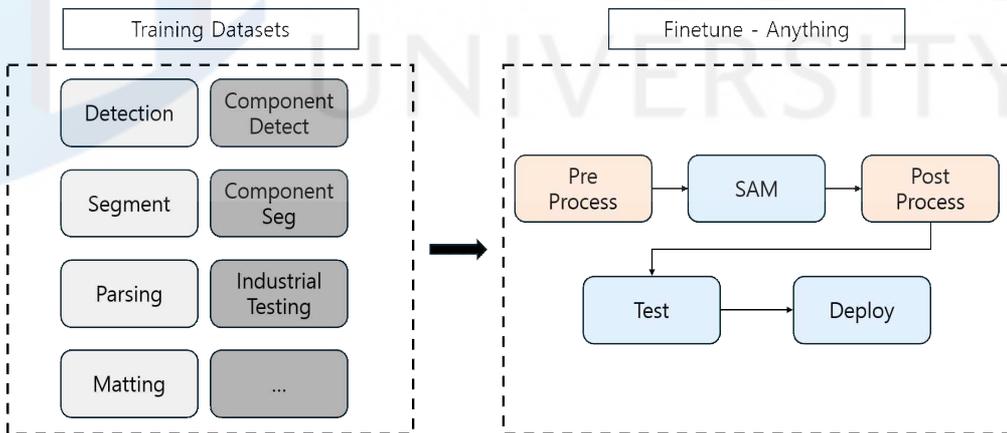
Fine-tuning 과정은 일반적으로 학습 데이터셋을 구축하고, 데이터셋 특성에 맞게 모델의 일부 가중치를 미세조정하여 성능을 개선한다. 이를 통해 SAM의 범용성은 유지하면서 특정 객체에 대한 정밀도를 향상시킬 수 있다. fine-tuning에 대한 개념은 [그림 3-4]에 나타내었다.

최근 다양한 분야에 활용하기 위해 SAM에 fine-tuning을 적용하는 연구가 활발하게 진행되고 있다. Peng et al. (2024)은 파라미터 공간을 재구성한

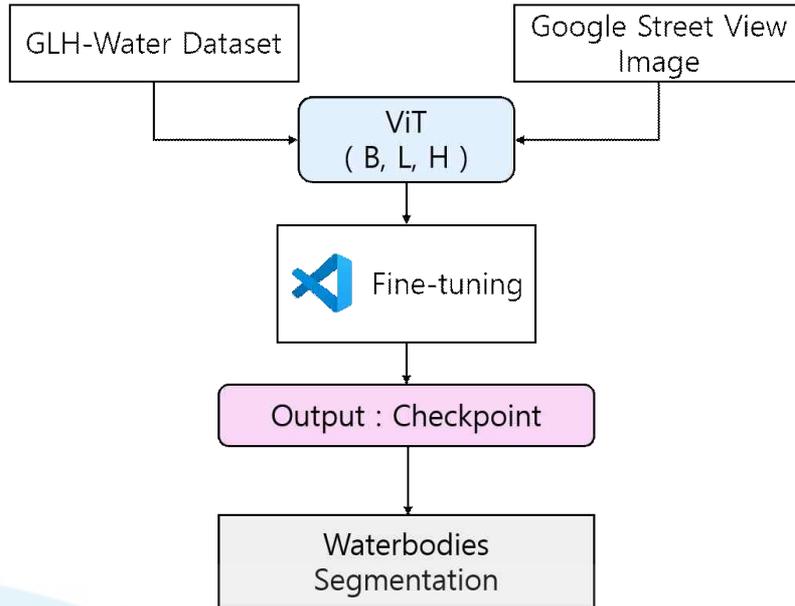
Sam-Parser 기법을 통해 적은 양의 학습 데이터로도 특정 도메인에서 성능 향상이 가능함을 보였다. 또한 Chai et al. (2024)은 Ladder Fine-tuning Approach을 사용하여 상호 보완적인 네트워크를 결합하였으며 이를 통해 SAM의 표현력을 강화하고 분할 정확도를 높이는 방법을 제안하였다. 이러한 연구들은 fine-tuning을 수행한 SAM이 SAM의 기본 모델에 비해 다양한 도메인에서 활용 가능하다는 것을 보여준다.

[그림 3-5]는 이번 연구에서 사용한 fine-tuning 방법으로, GLH-Water 데이터셋에 GSV에서 수집한 국내 수부 이미지를 추가하여 학습 데이터셋을 생성해 활용하였다. 이때 ViT 모델은 ViT-L 모델을 기반으로 학습하였으며, 학습 결과는 체크포인트(checkpoint) 파일로 저장하였다.

Fine-tuning은 특히 수채 탐지와 같은 고해상도 영상 분석에 있어 학습 데이터와 도메인 특화 문제의 한계를 극복할 수 있다는 점에서 주목받고 있다. 연구에서도 이러한 fine-tuning을 통해 SAM의 활용성을 확장하고자 하였다.



[그림 3-4] Fine-tuning 개념 및 흐름도



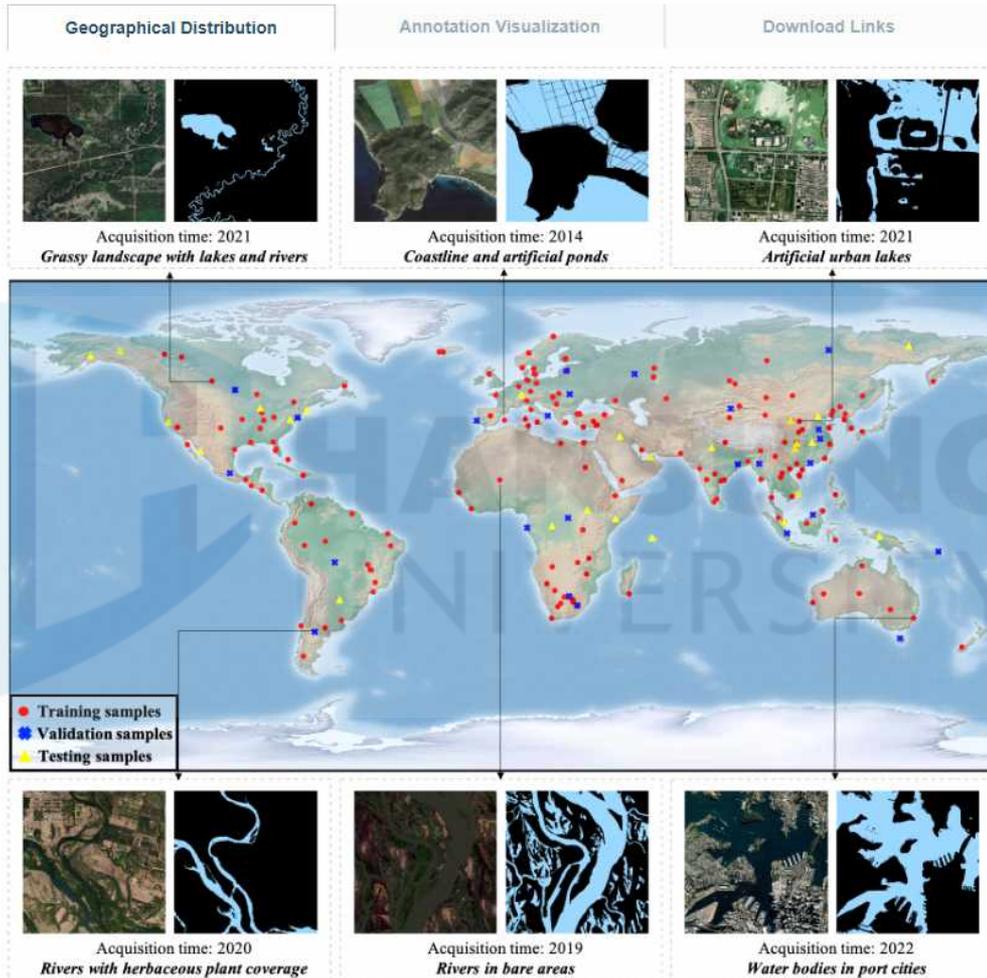
[그림 3-5] Fine-tuning 방법

1) GLH-Water 데이터셋

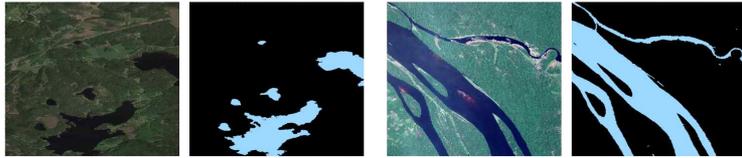
연구에서 학습 데이터셋으로 사용한 GLH-Water 데이터셋은 수체 관련 오픈소스 데이터셋으로 중국 우한 대학교 연구팀이 제작하였다. [그림 3-6]과 같이 전 세계를 포함하여 다양한 지역에서 수집된 250장의 초고해상도 (VHR: Very High Resolution) 위성 영상으로 구성되어 있으며, 약 40억 9천6백만 개의 픽셀 단위 수체 레이블이 포함되어 있다(Li et al., 2024). 각 영상은 해상도 0.3m GSD (Ground Sampling Distance)로 촬영되었으며, 크기는 12,800×12,800 픽셀로 거대한 데이터를 제공한다. 이러한 데이터셋은 다양한 지리와 환경으로부터 강, 호수, 저수지뿐만 아니라 산림, 관개지, 나지, 도시 지역 내 수체 등 매우 다양한 수체 유형을 포함하고 있다.

GLH-Water 데이터셋은 수작업으로 라벨링된 고정밀 수체 주석 데이터를 포함하고 있어, 학습 및 평가 과정에서 신뢰도 높은 Ground Truth 데이터를 제공한다. 라벨링 정보는 모델 학습과 성능 검증에 중요한 역할을 하며, [그림 3-7]는 데이터셋에 포함된 수체 라벨링 정보를 시각적으로 보여준

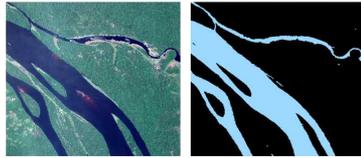
다. 하늘색 부분은 수체 부분이며, 검은색 부분은 비수체 부분이다. GLH-Water 데이터셋은 고해상도 수체 탐지 및 분할 연구에서 중요한 자료로, 기존 모델 성능 평가 및 새로운 분할 알고리즘 개발에 필수적인 역할을 한다.



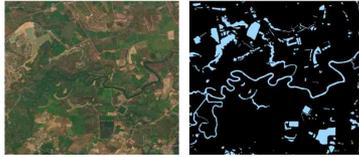
[그림 3-6] GLH-Water 데이터셋의 전 세계 위성 영상 분포도
<https://jack-bo1220.github.io/project/GLH-water.html>



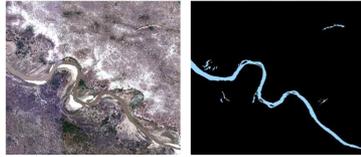
Lakes in the forest



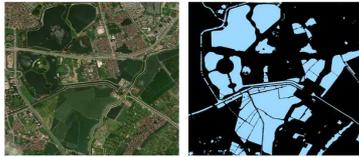
Rivers in the forest



Lakes and rivers in irrigated fields



Rivers in the shrub



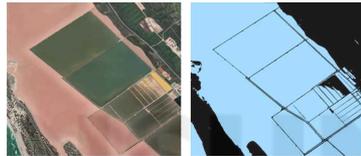
Lakes in the urban area



Rivers in the urban area



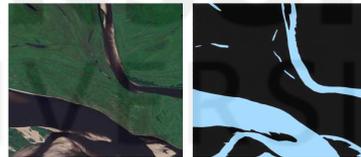
Glacial lakes



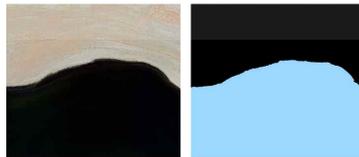
Lakes in the special scenario



Lakes in the grassland



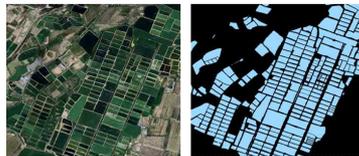
Rivers in the grassland



Lakes in the bare area



Rivers in the bare area



Artificial ponds



Water surrounding the port



Pure water body



Non-water body

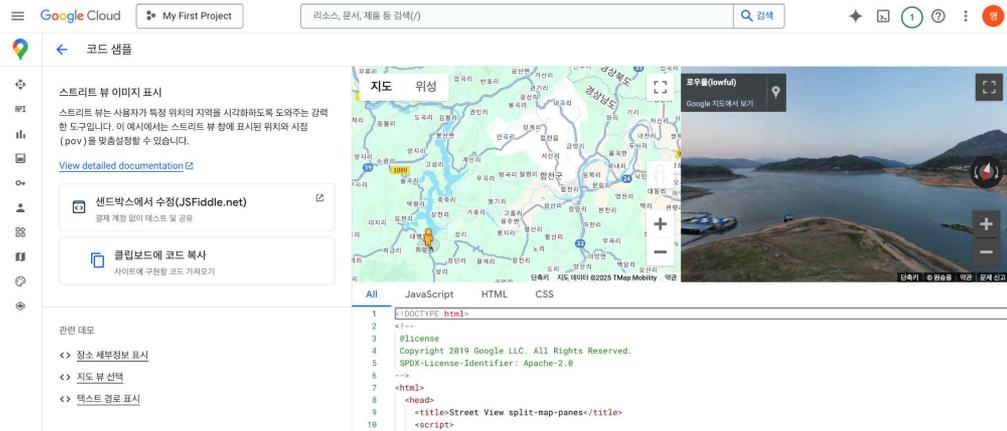
[그림 3-7] GLH-Water가 제공하는 라벨 예시

3) 학습 데이터셋 생성

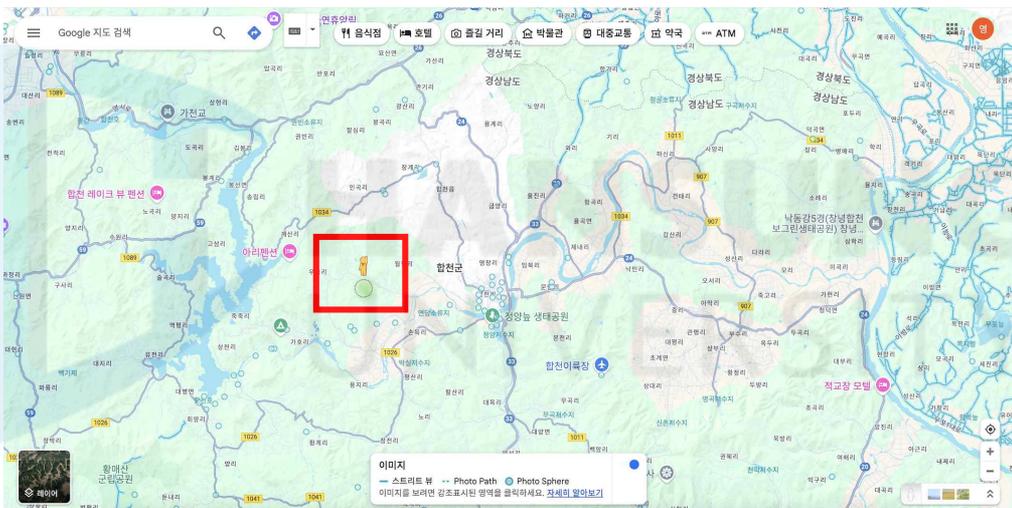
학습 데이터셋으로 GLH-Water 데이터셋의 단독 사용은 국내 연구 지역의 수체 환경을 충분히 반영하기 어려울 것으로 예상하였다. 따라서 GSV를 활용하여 국내 수체 이미지를 수집 후 데이터셋에 추가하였다. GSV는 기본적으로 Google에서 수집한 영상과 사용자가 업로드한 영상 두 가지로 구성된다. Google에서 제공하는 영상은 360도 카메라를 장착한 차량(street view car), 배낭(trekker), 배 등을 이용해서 영상을 촬영한다. 사용자가 업로드한 영상은 사용자가 직접 360도 카메라 또는 스마트폰으로 촬영한 이미지를 업로드하는 것으로 사용자 생성 콘텐츠(user-contributed content)에 추가된다.

GSV 데이터를 수집하기 위해서는 [그림 3-8]의 (a) Google Cloud Platform (GCP)에서 프로젝트 생성 후 Street View Static API를 활성화해야 한다. API Key를 발급받은 후 python 스크립트를 연동하여 국내 수부가 포함된 이미지를 다운로드하였다. 수집 방법은 위·경도와, (b) 구글맵에서 페그맨(Pegman)을 사용하였다. 페그맨은 구글맵에서 Street View 기능을 활성화하기 위한 도구로, 지도 위로 페그맨을 드래그하여 영상을 즉시 확인할 수 있다. 특히 페그맨을 사용하여 지도 위로 드래그하면 영상이 있는 곳은 지도상에 파란색 점으로 표시되기 때문에 관심 지역에 이미지가 존재하는지 사전 확인이 가능하다. 데이터들은 연구 지역 이외 국내 전체 지역을 포함하여 수집되었다.

이렇게 수집된 약 300개의 GSV 기반 수부 이미지들은 전처리 과정을 거쳐 학습 데이터셋에 포함하였다. 먼저 수집한 이미지의 크기를 모두 1024×1024 로 표준화하였다. SAM의 ViT 모델들은 입력 해상도를 모두 1024×1024 크기로 학습했기 때문에, 수집한 이미지도 동일한 크기로 생성하였다. 이후 [그림 3-9] (a)와 같이 하단에 작성된 Google 출처를 삭제한 뒤 (b) 이미지 내 수부를 확인하고 SAM을 사용하여 수체를 추출하였다. [그림 3-10]은 데이터셋에 추가한 GSV 기반 이미지들 중 일부를 시각화한 것으로 하늘색 부분이 수부 부분이다. 마지막으로 GLH-Water 데이터셋과 GSV 기반 이미지들을 병합하여 학습 데이터셋을 생성하였다.

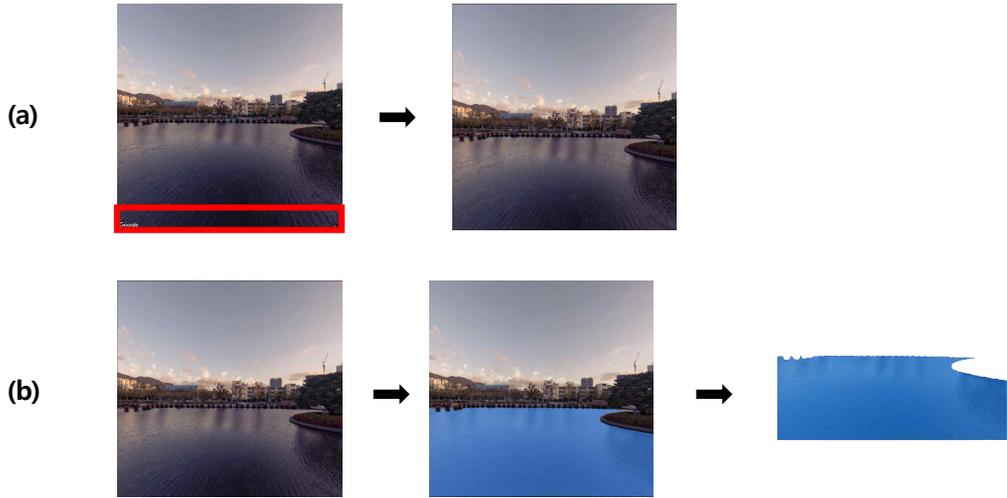


(a)



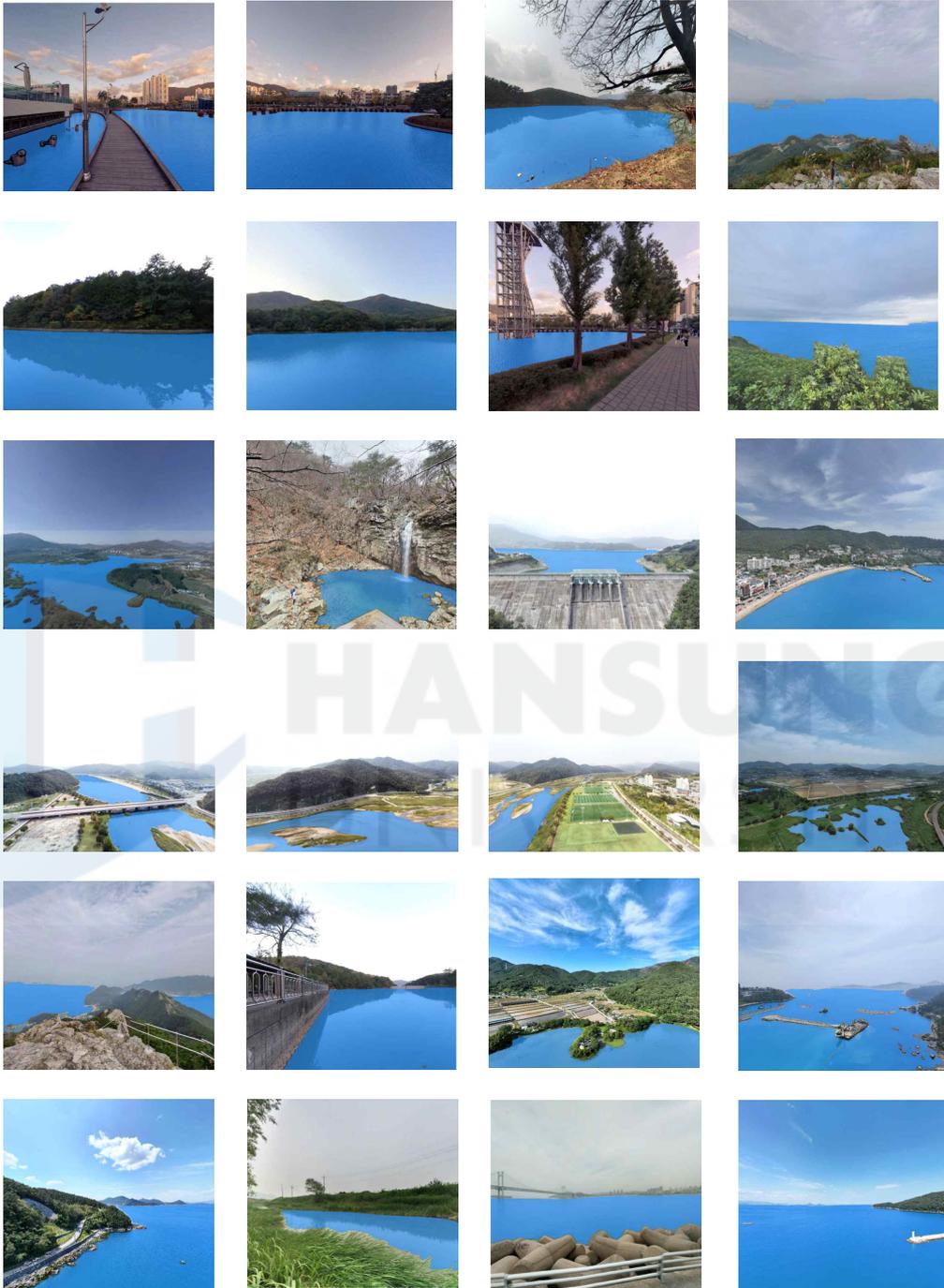
(b)

[그림 3-8] 학습 데이터셋 생성 방법
(a) Google Cloud Platform의 Street View 화면, (b) Google map 화면 속 페그맨



[그림 3-9] GSV 이미지 전처리 과정
 (a) 하단 Google 출처 삭제, (b) 이미지 내 수체 추출





[그림 3-10] 데이터셋에 추가한 GSV 이미지

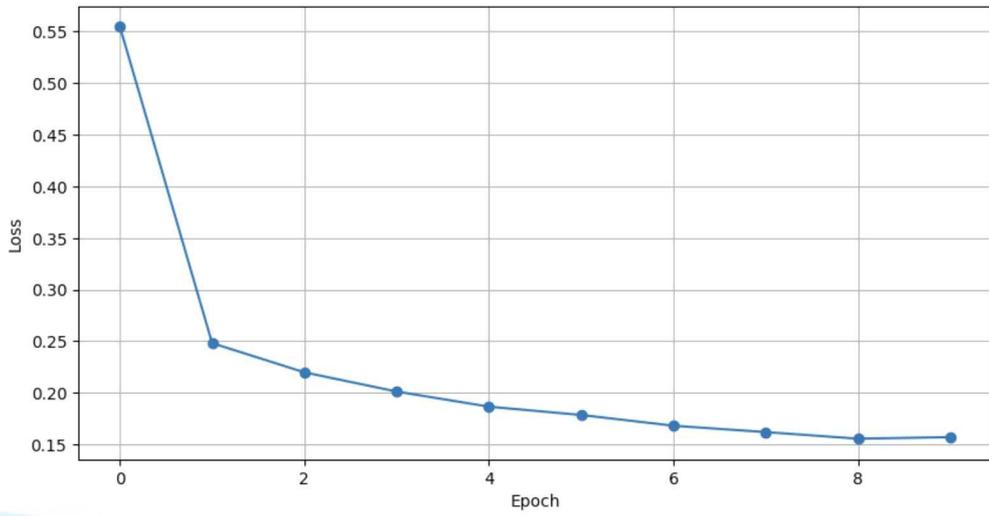
4) 데이터셋을 이용한 Fine-tuning

이번 연구에서는 부분적 학습(partial fine-tuning) 방법을 사용하여 SAM의 fine-tuning을 수행하였다. 부분적 학습은 전체 파라미터를 학습시키는 full fine-tuning과 달리, 특정 계층 또는 모듈만을 선택적으로 학습시켜 효율성과 성능을 모두 고려하는 방법이다(Shen et al., 2021). 이 방법은 특정 부분만을 학습시키지만 새로운 데이터셋에 대한 적응력은 향상시키고, 동시에 연산 비용과 메모리 사용량은 줄일 수 있다. 특히 Fu et al. (2023)은 전체가 아닌 적은 양의 학습 파라미터만 업데이트해도 full fine-tuning에 근접한 성능을 보일 수 있다고 발표하였다.

SAM은 사전 학습된 거대한 파라미터 공간을 보유하며, ViT를 기반으로 강력한 이미지 인코더와 분할 디코더를 포함한다. 전체 네트워크를 fine-tuning 하는 것은 계산량과 메모리 사용량을 매우 증가시키며, 작은 데이터셋은 오히려 과적합의 위험이 존재할 수 있다. 따라서 이미지 인코더는 고정하고 프롬프트 인코더와 마스크 디코더 부분만을 학습시키는 방법을 선택하였다.

[그림 3-11]은 학습 데이터셋의 fine-tuning 과정에서 나타난 학습 손실(training loss)을 에포크(epoch)별로 시각화한 결과를 보여준다. 초기 학습 단계에서 손실 값이 급격히 감소 후 학습의 안정화 과정을 진행하였으며, 후반부에는 완만한 감소로 학습이 안정화됨을 확인할 수 있었다. 이러한 모습은 학습이 새로운 데이터셋에 빠르게 적응하며 안정적으로 학습되었음을 보여준다.

결과적으로, 이번 연구의 부분적 학습 기법은 적은 학습 데이터와 제한된 자원에서도 효과적인 계산과 적응력을 가능하게 하였으며, SAM의 범용성과 세부 객체 탐지 성능을 동시에 확보할 수 있는 최적의 fine-tuning 전략임을 확인할 수 있었다.



[그림 3-11] Fine-tuning 학습 손실 그래프



HANSUNG
UNIVERSITY

제 5 절 수체 추출 실험

1) Geo-SAM 적용

Geo-SAM과 ViT-L 모델을 이용하여 수체 추출 실험을 진행하였으며, [그림 3-13]은 (a) KOMPSAT-3 영상과 (b) Wavelet 기반 융합 영상에 대해 Geo-SAM을 적용한 결과이다. 두 영상의 주요 수체가 모두 안정적으로 추출되었음을 확인할 수 있다. 대규모 수체인 합천호뿐만 아니라 산간 지역의 소규모 저수지까지도 분할이 가능하였다. 특히 Wavelet 기반 융합 영상의 경우 경계선의 형태가 KOMPSAT-3 영상 대비 뚜렷하게 나타났으며, [그림 3-12]와 같이 합천호에 설치된 매화 모양의 수상 태양광 패널을 관찰할 수 있었다. 하지만 물의 연속성이 부족하여 끊어져 보이는 지형에서는 분할 정확도에 한계가 있었다. 이러한 결과는 Geo-SAM이 기본적인 수부 추출 가능성을 보여주지만, 추가적인 fine-tuning 필요성을 제시한다.

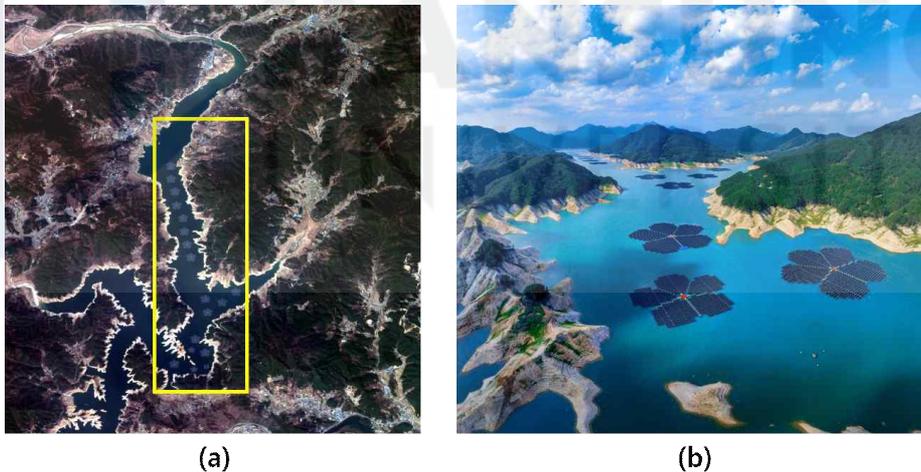
2) SAM 적용

[그림 3-14]은 fine-tuning을 진행하지 않고 SAM을 단독 사용하여 (a) KOMPSAT-3 영상과 (b) Wavelet 기반 융합 영상에 대해 수부 추출을 진행한 결과이다. Geo-SAM과 동일하게 별도의 데이터셋 학습 없이 ViT-L 모델을 사용하여 분할을 진행하였다. 두 영상 모두 합천호에 설치된 태양광 패널은 관찰할 수 있으나, KOMPSAT-3 영상의 경우 경계가 뚜렷하지 않고 원형 모양으로 분할되었다. 또한 두 영상 모두 작은 크기의 수체들이 제대로 추출되지 않아 분할의 누락이 존재한다.

이러한 결과는 fine-tuning을 진행하지 않은 SAM의 경우 수체와 같이 정형화되지 않은 특정 객체에서는 세부 정보를 반영하지 못해 분할의 정확성이 떨어질 수 있음을 의미한다. 또한 fine-tuning 적용 실험에서 SAM의 학습 데이터셋 및 미세조정을 통해 분할 성능이 얼마나 개선되는지를 평가하는 기초 자료로 활용될 수 있다.

3) Fine-tuning 적용

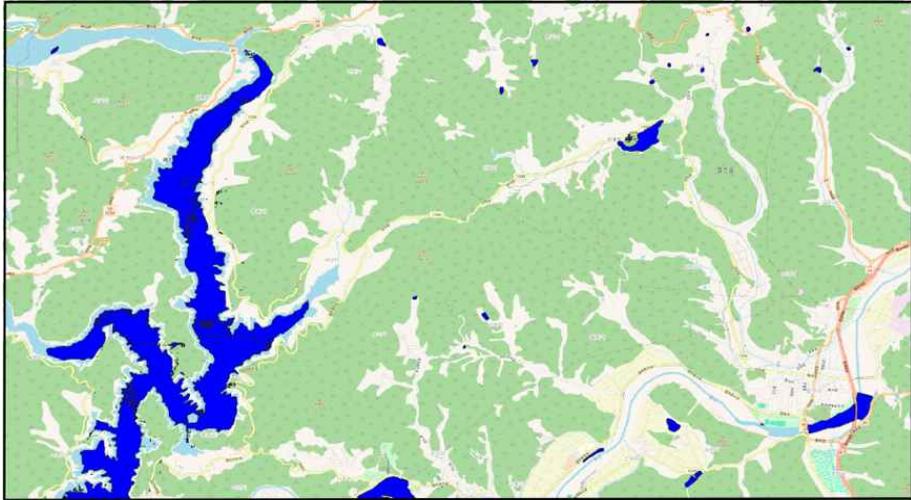
[그림 3-15]는 fine-tuned SAM을 사용하여 (a) KOMPSAT-3 영상과 (b) Wavelet 기반 융합 영상에 대해 수부 추출을 진행한 결과이다. Fine-tuned SAM은 Geo-SAM과 SAM에서 확인할 수 있었던 끊어진 모양의 수체가 대부분 연결된 형태로 추출되어 개선된 모습을 보였다. 특히 Wavelet 기반 융합 영상에서는 수체의 연속성뿐만 아니라 합천호에 설치된 태양광 패널이 위치대로 경계가 추출되었다. 이는 fine-tuning에 사용한 데이터셋을 통해 SAM이 특정 객체에 대해 효과적으로 학습했음을 나타낸다. 하지만 SAM과 마찬가지로 두 영상 모두 작은 크기의 수체는 제대로 추출되지 않아 분할의 누락이 존재한다. 이는 SAM과 fine-tuned SAM을 진행한 GUI의 한계로 생각된다.



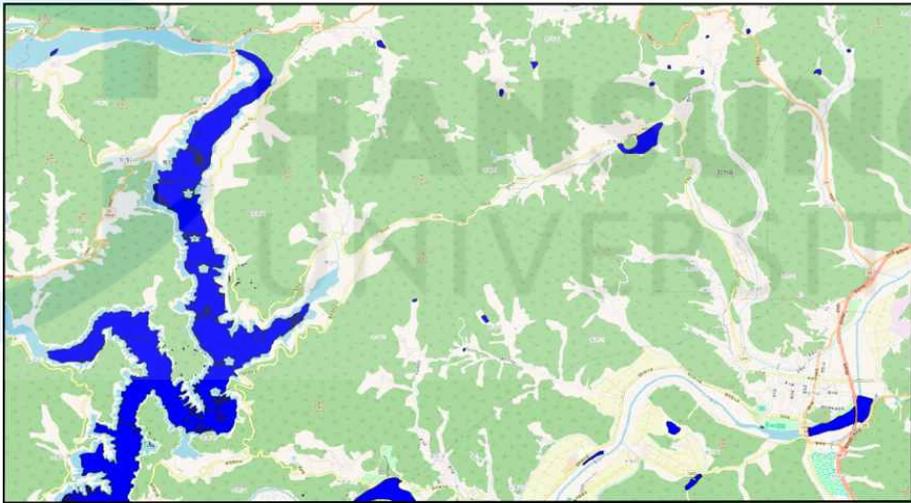
[그림 3-12] 합천호에 설치된 수상 태양광 패널

(a) 위성 영상에서 확인할 수 있는 태양광 패널, (b) 출처: 합천군청 공식

홈페이지(<https://www.hc.go.kr/06571/06779/06782.web?amode=view&idx=1200&>)

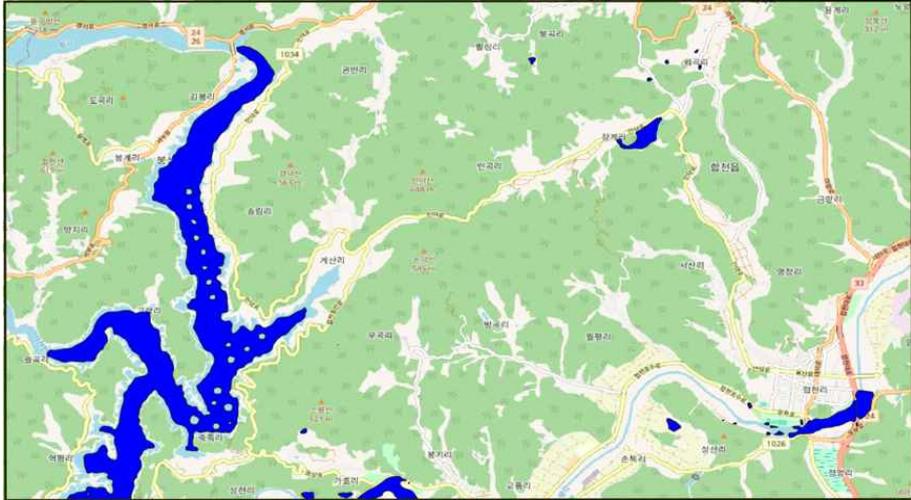


(a)

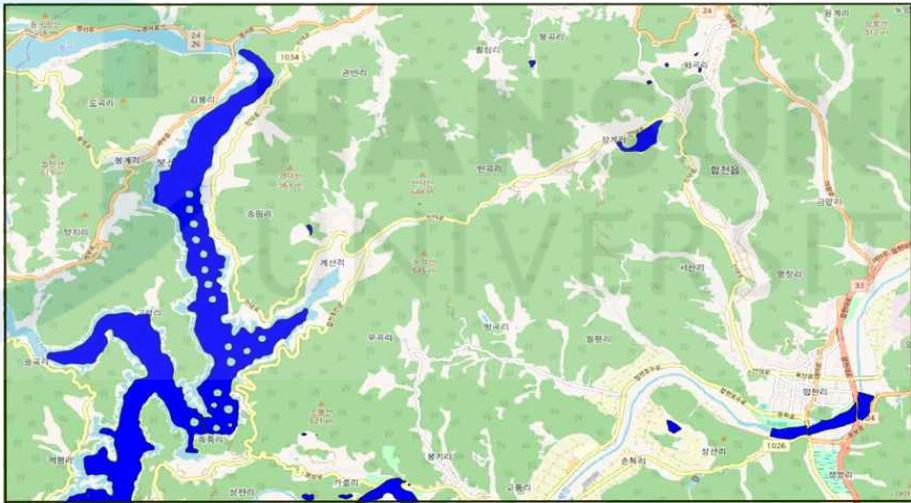


(b)

[그림 3-13] Geo-SAM 결과
(a) KOMPSAT-3 영상, (b) Wavelet 기반 융합 영상



(a)



(b)

[그림 3-15] Fine-tuning 수행 후 SAM의 결과
 (a) KOMPSAT-3 영상, (b) Wavelet 기반 융합 영상

제 4 장 수체 추출 결과 분석

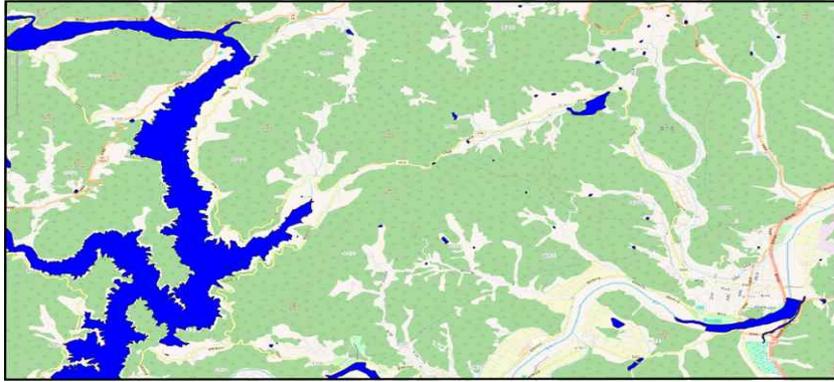
제 1 절 비교 자료

Geo-SAM, SAM, Fine-tuned SAM의 수체 추출 결과와 비교하기 위해 [그림 3-4]와 같이 비교 자료를 사용하였다.

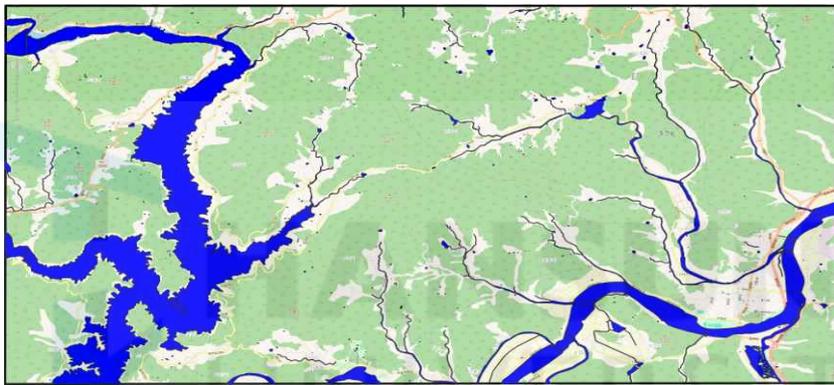
(a)는 QuickOSM으로, OSM 기반의 QGIS에서 사용 가능한 플러그인이다. 이 자료는 QuickOSM을 통해 연구 지역 범위에서 하천 및 수체 객체를 자동으로 추출한 것으로 넓은 범위의 수역을 나타낸다. 일반적으로 natural=water 태그를 사용하며 염전과 야외 인공 수영장 등은 포함되지 않는다.

(b)는 국토지리정보원(NGII: National Geographic Information Institute)에서 제공하는 수계 수치 자료이다. NGII의 수계로 분류된 자료 중, 하천 경계 자료와 호수/저수지 경계 자료를 병합하여 사용하였다. NGII 자료는 축척 1/5,000으로 제작되어 비교적 정확하고 세부적인 정보를 제공하지만, 갱신 주기 및 최신 데이터 여부에 따라 누락 또는 오차가 발생할 수 있다.

이번 연구에서 (c)는 실험 결과의 정확도를 평가하기 위한 Ground Truth 자료이다. 위성 영상의 촬영 시점과 비교 자료들의 제작 시점이 일치하지 않아 정확도가 감소할 수 있으므로, 변화 요인을 최소화하고 신뢰성 높은 자료를 준비하는 것은 중요하다. 신뢰할 수 있는 Ground Truth 자료를 확보하지 못하면 실험 결과의 정확성을 보장하기 어렵고, 모델 성능 평가 과정에서 잘못된 결론을 도출할 가능성이 높다. 따라서 KOMPSAT-3 영상과 네이버 지도, 카카오맵, OSM을 참조하여 직접 수작업 라벨링으로 생성하였다. Ground Truth 제작에 사용한 자료는 [표 4-1]에 정리하였다.



(a)



(b)



(c)

[그림 4-1] 비교 자료
(a) QuickOSM 기반 추출 수체 자료, (b) NGII 수체 자료,
(c) Ground Truth

[표 4-1] Ground Truth 생성 자료

자료	정보	취득 날짜
KOMPSAT-3 영상	[표 3-1] KOMPSAT-3 영상과 동일	2023.03.10
네이버 지도	최신 지도만 제공, 과거 날짜의 지도 조회 불가	버전 정보 v3.21.0
카카오맵	- 네이버 지도와 동일하게 과거 데이터 조회 불가 - QGIS 플러그인 TMS for Korea 사용	버전 정보 v3.0.5
OpenStreetMap	- 단순한 지도라기보다 지리공간 데이터베이스 - OpenStreetMap Korea 커뮤니티 존재 - 이미지뿐만 아니라 데이터베이스 제공	QGIS v3.12.1

제 2 절 성능 평가 지표

객체 탐지 및 분할 성능을 평가하기 위해 다양한 지표가 사용된다. 대표적으로 정밀도(Precision), 재현율(Recall), F1-Score, Accuracy, 그리고 Intersection over Union (IoU) 등이 있다. 정밀도와 재현율은 객체 탐지의 참과 거짓을 기준으로 판단하여 전반적인 탐지율과 정확도를 보여주지만, 객체의 모양과 위치의 정확성까지 반영하기는 어렵다(Chen et al., 2017). Accuracy는 전체 분류 결과의 비율만을 계산하므로 객체의 위치 일관성을 직접적으로 평가하지는 않는다. 반면, IoU는 예측한 객체와 실제 객체의 교차 영역과 합집합 영역을 비교하여 계산되기 때문에 분할된 객체의 형태 및 위치 정확도를 평가하는 데 유리하다.

따라서 연구에서 선택한 성능 평가 지표는 IoU이며, 다양한 형태의 수체를 정확히 평가하기 위해 IoU의 확장 개념인 mIoU와 WIoU를 활용하였다.

1) mean Intersection over Union (mIoU)

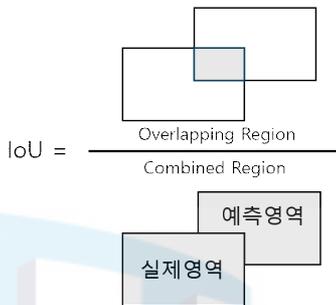
mIoU는 객체 분할 성능을 평가하는 대표적인 지표 중 하나로, IoU 값의 평균값이다. IoU의 평균은 먼저 개별 클래스 또는 객체마다 교차 영역의 값(IoU)을 계산하고, 이를 모두 더해 클래스 또는 객체의 수로 나누어 계산한다. 단일 클래스에 대해 성능을 평가한다면 IoU 값을 사용하면 되지만, 다중 클래스에 대해 성능을 평가하고 싶다면 mIoU 값을 사용하는 것이 좋다. mIoU는 분할된 객체의 위치와 모양의 정확도를 반영하고 동시에 전반적인 일치도를 직관적으로 파악할 수 있다는 장점이 있다.

[그림 3-15]는 IoU의 개념과 mIoU 계산 과정을 시각화한 것이다. 실제 영역과 예측 영역 사이의 교집합과 합집합 영역의 비율을 계산하고, 이들의 평균값을 계산하여 전체 mIoU를 산출하는 과정을 보여준다. Behera et al. (2021)의 Faster R-CNN 기반의 과일 수확량 추정 연구에서 정확도 평가의 핵심 지표로 mIoU를 사용하여 객체별 경계 정밀도와 예측 위치의 일치도를 동시에 검증하였다. 이처럼 mIoU는 다양한 크기와 형태 그리고 위치 정합성을 고려하였을 때, 이번 연구의 수채 추출 결과를 평가할 수 있는 적합한 지표이다.

2) Weighted Intersection over Union (WIoU)

mIoU는 모든 객체를 동일한 비중으로 고려하기 때문에, 객체 간 크기 차이가 반영되지 않는다. 이에 Tetteh et al. (2021)은 객체의 크기에 따라 가중치를 부여하고, 큰 객체일수록 분할 평가에 더 큰 영향을 주도록 WIoU를 제안하였다. 연구에서는 이를 참고하여, 크기가 큰 수채에 더 큰 가중치를 부여하는 방법을 적용하였다. 예를 들어 [그림 3-16]의 a, b, c, d는 동일한 위치의 객체이며 크기는 $a > b > c > d$ 순으로 면적이 크다. 따라서 면적이 큰 순서대로 더 높은 가중치를 차등 적용하여 큰 객체의 분할 정확도는 전체 성능 평가에 반영되고 작은 객체는 전체 성능에 미치는 영향을 줄이도록 하였다.

WIoU는 넓은 면적의 주요 수체가 평가 결과에 더 크게 반영되고, 상대적으로 작은 객체가 과도한 영향을 미치지 않도록 설계되었다. 이를 통해, 크기 변화가 큰 수체의 분할 평가에서 신뢰도 높은 결과를 얻을 수 있다.



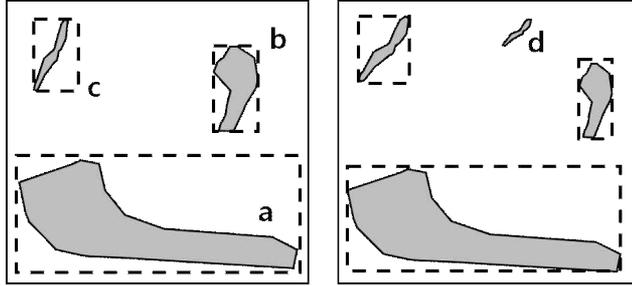
$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

- A : 실제 영역
- B : 예측 영역
- $|A \cap B|$: 두 영역의 교집합(Intersection)
- $|A \cup B|$: 두 영역의 합집합(Union)

$$mIoU = \sum_{i=1}^N \frac{|A_i \cap B_i|}{|A_i \cup B_i|}$$

- A_i : i 번째 클래스의 Ground Truth
- B_i : i 번째 클래스의 예측된 영역
- $|A_i \cap B_i|$: i 번째 클래스에서 교집합 영역
- $|A_i \cup B_i|$: i 번째 클래스에서 합집합 영역

[그림 4-2] IoU와 mIoU의 개념 및 계산식



1. 각 객체에 대한 IoU 계산

$$IoU(x) = \frac{Area\ of\ Intersection(x)}{Area\ of\ Union(x)}$$

2. 모든 객체가 겹치는 영역(교차 영역)을 계산하고 다음 값을 합산

$$\begin{aligned} Total\ Intersection\ Area \\ = Area\ of\ Intersection(a) + Area\ of\ Intersection(b) \\ + Area\ of\ Intersection(c) \end{aligned}$$

3. 각 객체에 대해 개별 교차 영역을 전체 교차 영역으로 나누어 가중치를 구함

$$Weight(x) = \frac{Area\ of\ Intersection(x)}{Total\ Intersection\ Area}$$

4. 각 객체의 IoU를 가중치로 곱하여 가중치가 적용된 IoU를 구함

$$WIoU = IoU(x) \times Weight(x)$$

[그림 4-3] WIoU 개념 및 계산식

[표 4-2] KOMPSAT-3 광학 영상에 대한 계산 결과

WIoU mIoU	SAM	Fine-tuned SAM	Geo-SAM	Quick OSM	NGII	Ground Truth
SAM	1.0000	0.8982	0.8717	0.6600	0.6447	0.8570
Fine-tuned SAM	0.8744	1.0000	0.8719	0.6957	0.6839	0.9420
Geo-SAM	0.8531	0.8629	1.0000	0.7172	0.6847	0.9191
Quick OSM	0.6131	0.6529	0.6882	1.0000	0.8746	0.7292
NGII	0.5609	0.6111	0.6167	0.8167	1.0000	0.7112
Ground Truth	0.8190	0.8843	0.8591	0.6919	0.6373	1.0000

[표 4-3] Wavelet 기반 융합 영상에 대한 계산 결과

WIoU mIoU	SAM	Fine-tuned SAM	Geo-SAM	Quick OSM	NGII	Ground Truth
SAM	1.0000	0.9074	0.8893	0.6599	0.6462	0.8540
Fine-tuned SAM	0.8818	1.0000	0.8956	0.7013	0.6889	0.9562
Geo-SAM	0.8725	0.8693	1.0000	0.7121	0.6845	0.9295
Quick OSM	0.6143	0.6566	0.6957	1.0000	0.8746	0.7292
NGII	0.5627	0.6150	0.6177	0.8167	1.0000	0.7112
Ground Truth	0.8213	0.9119	0.8704	0.6919	0.6373	1.0000

제 3 절 결과 분석

KOMPSAT-3 광학 영상과 Wavelet 기반 융합 영상을 대상으로, Geo-SAM, SAM, Fine-tuned SAM을 사용하여 수체 추출을 진행하고, 추출된 결과를 QuickOSM 수체 자료와 NGII 수계 수치 자료, Ground Truth와 비교하였다. [표 4-1]과 [표 4-2]는 각각 KOMPSAT-3 영상과 Wavelet 기반 융합 영상을 mIoU와 WIoU 지표를 사용하여 비교한 결과이다.

[표 4-1] KOMPSAT-3 영상에 대한 실험 결과에서, Geo-SAM 기준 fine-tuned SAM의 mIoU 0.8843, WIoU 0.9420으로 가장 높은 성능을 보였다. SAM과 Geo-SAM 또한 각각 mIoU 0.8190, 0.8591, WIoU 0.8570, 0.8717로 높은 성능을 보였으나 fine-tuned SAM 보다 다소 낮은 성능을 나타내었다. 반면, QuickOSM과 NGII는 mIoU 0.6919, 0.6373 WIoU 0.7292, 0.7112로 상대적으로 낮은 정확도를 보였다.

[표 4-2] Wavelet 기반 융합 영상에 대한 실험 결과에서도, Geo-SAM 기준 Fine-tuned SAM의 결과가 0.9119, 0.9562로 가장 높은 성능을 보였다. SAM과 Geo-SAM 또한 mIoU 0.8213, 0.8704 WIoU 0.8540, 0.9295로 KOMPSAT-3 영상 실험과 비슷하게 결과가 나타났다. QuickOSM과 NGII의 mIoU와 WIoU 또한 모두 낮은 정확도를 보였다.

Fine-tuned SAM이 KOMPSAT-3 영상과 Wavelet 기반 융합 영상과 비교했을 때 가장 우수한 분할 성능을 나타냈다. 이는 GLH-Water 데이터셋과 GSV 이미지를 활용한 미세조정을 통해 다양한 크기와 모양의 수체를 보다 효과적으로 탐지할 수 있었음을 의미한다. 반면 QuickOSM과 NGII는 상대적으로 낮은 정확도를 나타냈는데, 이는 위성 영상의 촬영 시점과 자료의 생성 및 갱신 시점의 차이로 인한 것으로 판단된다.

제 5 장 결 론

수부의 변화를 신속하고 정확하게 파악하는 것은 수자원 관리, 농업 및 환경 모니터링, 재난 관리 등 다양한 분야에 핵심적인 과제이다. 하지만 수부는 크기와 형태가 매우 다양하며, 계절 변화와 강우량과 같은 요인으로 인해 지속적으로 모양이 바뀔 수 있다. 따라서 수부를 안정적으로 탐지하고 분할할 수 있는 기술 개발이 필수적이다.

최근 고해상도 위성 영상의 발전과 함께 융합 영상의 활용이 주목받고 있다. 광학 영상은 직관적인 분석이 가능하지만 대기 조건과 일조량에 제약이 있다. 반면 SAR 영상은 기상 조건에 관계없이 안정적인 촬영이 가능하지만 노이즈와 스펙클 현상으로 경계가 명확하지 않을 수 있다. 이러한 장단점을 상호 보완하기 위해 Wavelet 기반 이미지 융합 기법을 사용하여 보다 정밀한 수체 추출을 가능하게 하였다.

그러나 융합 영상만으로는 수부를 탐지하는 데 한계가 존재한다. 이에 Meta AI에서 발표한 대규모 비전 모델인 SAM을 사용하여 수부를 추출하였다. SAM은 대규모 SA-1B 데이터셋을 통해 학습되어 다양한 도메인에서 범용적인 분할 성능을 확보하였다. 또한 제로 샷 분할이 가능하다는 점에서 기존 알고리즘 기반 및 딥러닝 기반 분할과 차별화된다. 그러나 SAM의 기본 모델만으로는 복잡한 경계나 다양한 모양 및 크기를 가진 수체와 같은 객체를 탐지하기에는 한계가 존재한다.

이를 해결하기 위해 GLH-Water 데이터셋에 GSV 기반 국내 수부 이미지를 추가하여 수체 학습용 데이터셋을 구축하였다. 이후 fine-tuning을 통해 수체 탐지에 적합한 모델을 생성하였다.

KOMPSAT-3 광학 영상과 Wavelet 기반 융합 영상으로 SAM, Fine-tuned SAM 그리고 Geo-SAM을 사용하여 수체를 추출하였으며 QuickOSM 수체 자료와 NGII 수계 수치 자료, 그리고 직접 제작한 Ground Truth 자료와 함께 mIoU 및 WIoU 지표를 사용하여 결과를 평가하였다. 그 결과, Ground Truth 기준으로 KOMPSAT-3 영상의 평균

mIoU가 0.7783, WIoU는 0.8317이었으며, Wavelet 기반 융합 영상의 평균 mIoU가 0.7866, WIoU는 0.8360으로 광학 영상보다 융합 영상에서 더 높은 성능을 보였다. 이는 융합 영상이 복잡한 환경에서도 안정적이고 정밀한 수부 탐지를 가능하게 하기 때문이다.

이번 연구에서는 KOMPSAT-3 광학 영상, Wavelet 기반의 융합 영상, 그리고 SAM의 fine-tuning 기법을 결합하여 수체를 안정적으로 추출하는 방법을 제안하였다. 이를 위해 GLH-Water 데이터셋과 국내 GSV 수부 이미지를 활용하여 학습 데이터셋을 구축하고, SAM의 디코더를 중심으로 재학습을 수행하였다. 그 결과, fine-tuned SAM이 기존 모델에 비해 복잡한 경계를 가진 수체를 보다 정밀하게 탐지하고, 연속된 형태로 추출할 수 있음을 확인하였다. 또한, 단일 광학 영상보다 Wavelet 기반 융합 영상을 활용했을 때 수체 추출 정확도가 향상되는 것으로 나타났다. 반면 QuickOSM과 NGII 자료는 위성 영상 촬영 시점과 데이터 제작 시점의 차이로 인해 낮은 성능을 보였다. 이를 통해 Ground Truth 자료의 최신성과 정확성이 연구 결과에 영향을 미친다는 점을 확인할 수 있었으며, 향후 연구에서는 최신 데이터를 기반으로 한 신뢰도 높은 Ground Truth 자료의 확보가 중요하다는 점이 강조된다. 연구 결과는 환경 모니터링, 홍수 예측 등 재해 대응 시스템에서 수체를 실시간으로 파악하고 관리하는 데 활용될 수 있을 것으로 기대된다.

참 고 문 헌

1. 국내문헌

조나혜, 이정주, 김현덕. (2023). 고해상도 위성 영상과 인공지능을 활용한 국토 변화탐지 및 모니터링 연구: 실증대상 지역인 정읍시를 중심으로. *지적과 국토정보*, 53(1), 107-121.

2. 국외문헌

An, S., & Rui, X. (2022). A high-precision water body extraction method based on improved lightweight U-Net. *Remote Sensing*, 14(17), 4127.

Byun, Y., Choi, J., & Han, Y. (2013). An area-based image fusion scheme for the integration of SAR and optical satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(5), 2212-2220.

Carraro, A., Sozzi, M., & Marinello, F. (2023). The Segment Anything Model (SAM) for accelerating the smart farming revolution. *Smart Agricultural Technology*, 6, 100367. <https://doi.org/10.1016/j.atech.2023.100367>

Chai, S., Jain, R. K., Teng, S., Liu, J., Li, Y., Tateyama, T., & Chen, Y. W. (2023). Ladder fine-tuning approach for SAM integrating complementary network. *arXiv preprint arXiv:2306.12737*.

Chamundeewari, V. V., Singh, D., & Singh, K. (2009). An analysis of texture measures in PCA-based unsupervised classification of SAR images. *IEEE Geoscience and Remote Sensing Letters*, 6(2), 214-

218.

- Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., ... & Qiao, Y. (2023). Sam-Med2D. arXiv preprint arXiv:2308.16184.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16×16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
- Feng, W., Zhu, L., & Yu, L. (2023). Cheap lunch for medical image segmentation by fine-tuning SAM on few exemplars. arXiv preprint arXiv:2308.14133.
- Fu, Z., Yang, H., So, A. M.-C., Lam, W., Bing, L., & Collier, N. (2023). On the effectiveness of parameter-efficient fine-tuning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(11), 12799–12807.
- Giannakis, I., Bhardwaj, A., Sam, L., & Leontidis, G. (2024). A flexible deep learning crater detection scheme using Segment Anything Model (SAM). *Icarus*, 408, 115797.
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2016). Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1), 142–158. <https://doi.org/10.1109/TPAMI.2015.2437384>
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580–587.
- Gui, S., Song, S., Qin, R., & Tang, Y. (2024). Remote sensing object detection in the deep learning era—A review. *Remote Sensing*, 16(2), 327.

- Hellweger, F. L., Schlosser, P., Lall, U., & Weissel, J. K. (2004). Use of satellite imagery for water quality studies in New York Harbor. *Estuarine, Coastal and Shelf Science*, 61(3), 437–448. <https://doi.org/10.1016/j.ecss.2004.06.019>
- Kakde, B. (2017). Comparative study of image fusion methods: A review. *SSRN Electronic Journal*.
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al. (2023). Segment anything. arXiv preprint arXiv:2304.02643. <https://doi.org/10.48550/arXiv.2304.02643>
- Kong, Y., Hong, F., Leung, H., & Peng, X. (2021). A fusion method of optical image and SAR image based on Dense-UGAN and Gram-Schmidt transformation. *Remote Sensing*, 13(21), 4274. <https://doi.org/10.3390/rs13214274>
- Kulkarni, S. C., & Rege, P. P. (2020). Pixel level fusion techniques for SAR and optical images: A review. *Information Fusion*, 59, 13–29.
- Li, H., Manjunath, B. S., & Mitra, S. K. (1995). Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3), 235–245.
- Li, Y., Dang, B., Li, W., & Zhang, Y. (2024). GLH-Water: A large-scale dataset for global surface water detection in large-size very-high-resolution satellite imagery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20), 22213–22221.
- Liu, C., Sun, Y., Xu, Y., Sun, Z., Zhang, X., Lei, L., & Kuang, G. (2024). A review of optical and SAR image deep feature fusion in semantic segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 12910–12930.
- Mazurowski, M. A., Dong, H., Gu, H., Yang, J., Konz, N., & Zhang,

- Y. (2023). Segment anything model for medical image analysis: An experimental study. *Medical Image Analysis*, 89, 102918. <https://doi.org/10.1016/j.media.2023.102918>
- Moghimi, A., Welzel, M., Celik, T., & Schlurmann, T. (2024). A comparative performance analysis of popular deep learning models and Segment Anything Model (SAM) for river water segmentation in close-range remote sensing imagery. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3385425>
- Oscó, L. P., Wu, Q., De Lemos, E. L., Gonçalves, W. N., Ramos, A. P. M., Li, J., & Junior, J. M. (2023). The Segment Anything Model (SAM) for remote sensing applications: From zero to one shot. *International Journal of Applied Earth Observation and Geoinformation*, 124, 103540.
- Pietrucha-Urbanik, K., & Rak, J. (2023). Water, resources, and resilience: Insights from diverse environmental studies. *Water*, 15(22), 3965. <https://doi.org/10.3390/w15223965>
- Ren, S., Luzzi, F., Lahrichi, S., Kassaw, K., Collins, L. M., Bradbury, K., & Malof, J. M. (2024). Segment anything, from space? *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 8355–8365.
- Ren, Y., Yang, X., Wang, Z., Yu, G., Liu, Y., Liu, X., Meng, D., Zhang, Q., & Yu, G. (2023). Segment Anything Model (SAM) assisted remote sensing supervision for mariculture—Using Liaoning Province, China as an example. *Remote Sensing*, 15(24), 5781.
- Shakya, A., Biswas, M., & Pal, M. (2020). CNN-based fusion and classification of SAR and optical data. *International Journal of Remote Sensing*, 41(22), 8839–8861.
- Shen, Z., Liu, Z., Qin, J., Savvides, M., & Cheng, K.-T. (2021). Partial

- is better than all: Revisiting fine-tuning strategy for few-shot learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(11), 9594–9602. <https://doi.org/10.1609/aaai.v35i11.17155>
- Tetteh, G. O., Gocht, A., Erasmi, S., Schwieder, M., & Conrad, C. (2021). Evaluation of Sentinel-1 and Sentinel-2 feature sets for delineating agricultural fields in heterogeneous landscapes. *IEEE Access*, 9, 116702–116719. <https://doi.org/10.1109/ACCESS.2021.3105903>
- Wang, X. (2016). Deep learning in object recognition, detection, and segmentation. *Foundations and Trends® in Signal Processing*, 8(4), 217–382. <http://dx.doi.org/10.1561/20000000071>
- Wu, Q., & Osco, L. P. (2023). samgeo: A Python package for segmenting geospatial data with the Segment Anything Model (SAM). *Journal of Open Source Software*, 8(89), 5663.
- Xia, Z., Li, Y., Guo, S., Chen, R., Zhang, W., Zhang, P., & Du, P. (2023). Mapping global water-surface photovoltaics with satellite images. *Renewable and Sustainable Energy Reviews*, 187, 113760. <https://doi.org/10.1016/j.rser.2023.113760>
- Xue, B., Cheng, H., Yang, Q., Wang, Y., & He, X. (2024). Adapting Segment Anything Model to aerial land cover classification with low-rank adaptation. *IEEE Geoscience and Remote Sensing Letters*, 21, 1–5.
- Yan, Z., Li, J., Li, X., Zhou, R., Zhang, W., Feng, Y., ... & Sun, X. (2023). RingMo-SAM: A foundation model for segment anything in multimodal remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–16. <https://doi.org/10.1109/TGRS.2023.3332219>
- Yang, R., He, G., Yin, R., Wang, G., Zhang, Z., Long, T., ... & Wang,

- J. (2024). A novel weakly-supervised method based on the segment anything model for seamless transition from classification to segmentation: A case study in segmenting latent photovoltaic locations. *International Journal of Applied Earth Observation and Geoinformation*, 130, 103929. <https://doi.org/10.1016/j.jag.2024.103929>
- Yehia, A., Elhifnawy, H., & Safy, M. (2019, December). An improved integrated intensity-hue-saturation with stationary wavelet transform multi-sensor image fusion approach. In *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 153–159.
- Zhang, C., Marfatia, P., Farhan, H., Di, L., Lin, L., Zhao, H., ... & Yang, Z. (2023, July). Enhancing USDA NASS cropland data layer with segment anything model. In *2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1–5.
- Zhao, Z., Fan, C., & Liu, L. (2023). Geo SAM: A QGIS plugin using segment anything model (SAM) to accelerate geospatial image segmentation. Zenodo.
- Zhou, T., Xia, W., Zhang, F., Chang, B., Wang, W., Yuan, Y., ... & Cremers, D. (2024). Image segmentation in foundation model era: A survey. *arXiv preprint arXiv:2408.12957*. <https://doi.org/10.48550/arXiv.2408.12957>

ABSTRACT

Object Extraction from Fused Optical and SAR Satellite Imagery Using the Segment Anything Model (SAM)

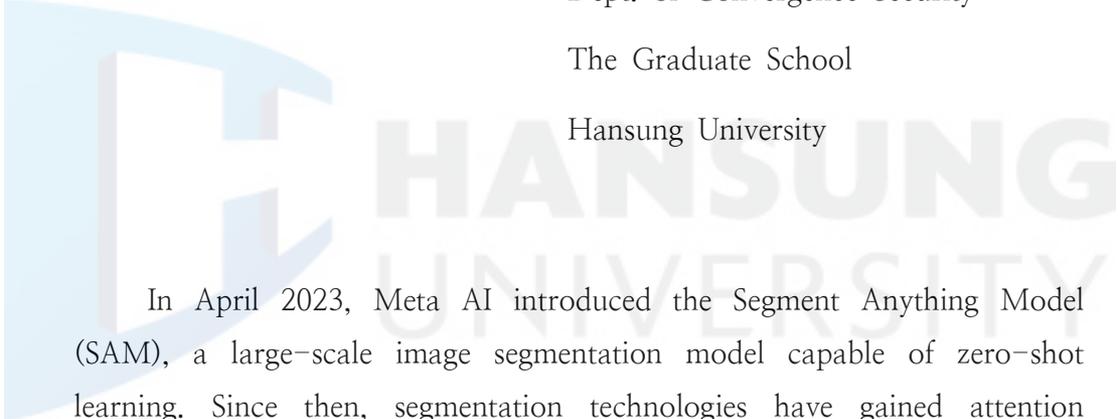
Lee, Ha-Young

Major in Convergence Security

Dept. of Convergence Security

The Graduate School

Hansung University



In April 2023, Meta AI introduced the Segment Anything Model (SAM), a large-scale image segmentation model capable of zero-shot learning. Since then, segmentation technologies have gained attention across various fields. In remote sensing, extracting water bodies from satellite images has become an important research topic due to its applications in environmental monitoring, disaster management, water resource assessment, and urban or agricultural planning. This study applies SAM to high-resolution satellite images by fusing KOMPSAT-3 optical images with KOMPSAT-5 Synthetic Aperture Radar (SAR) data to extract water bodies. Optical images offer color information but are affected by atmospheric conditions such as clouds, while SAR images are robust to weather but contain noise and lack detailed textures. To overcome these limitations, a wavelet-based image fusion method was

used to combine the complementary characteristics of both image types. SAM was applied to both the optical and fused images, and performance was evaluated using river maps from the National Geographic Information Institute (NGII) and OpenStreetMap (OSM) water data. Accuracy was measured with mean Intersection over Union (mIoU) and Weighted IoU (WIoU), showing improved results for the fused images. Although SAM performs well in general object segmentation, its baseline version is limited when detecting complex and irregular water shapes. To enhance its performance, fine-tuning was conducted using the GLH-Water dataset, supplemented by domestic water body images collected from Google Street View (GSV). The fine-tuned model achieved better accuracy in segmenting diverse and small-scale water features. These results demonstrate that water body detection in high-resolution satellite imagery can be improved through image fusion and SAM fine-tuning. The proposed method can be practically applied in fields that require reliable spatial information, such as disaster response and water resource management, by enabling stable segmentation of water bodies under various geographic and environmental conditions.

【Key words】 KOMPSAT-3, KOMPSAT-5, Segment Anything Model, Water Body, Wavelet-based Image Fusion, Fine-tuning