

복부 CT(Computed Tomography) 이미지 객체 분할 작업 성능 개선에 관한 연구

진태현, 장기태, 김두상

동서울대학교

honey7844@naver.com, 2470024@du.ac.kr, dskim@du.ac.kr

Research on Performance Improvement for Object Segmentation of Abdominal CT(Computed Tomography) Images

Jin Tae Heon, Jang Gi Tae, Kim Doo Sang

Dongseoul Univ.

요 약

본 논문은 HQ-SAM(Segment Anything in High Quality) 모델을 적용하여 복부 CT(Computed Tomography) 이미지 객체 분할(Segmentation)의 성능 개선 방식을 제안하였다. 기존 SAM(Segment Anything Model) 모델의 마스크 디코더에서, 고해상도 아웃풋 토큰(HQ-Output Token)과 글로벌-로컬 퓨전(Global-local Fusion)을 도입한 HQ-SAM 모델을 MICCAI FLARE22 Challenge Dataset을 통해 미세 조정(Fine-Tuning)을 하여 성능 개선을 이루었다. 본 논문에서 제안한 방식의 성능 검증을 위해 기존의 SAM, U-Net, DeepLabV3+ 모델을 같은 방식으로 미세 조정하여, 복부 CT 이미지 객체 분할한 결과를 Dice Score 측정 방식을 통해 비교 평가하였다. 실험 결과 본 논문에서 제안한 HQ-SAM 모델을 적용한 방식이 SAM, U-Net, DeepLabV3+ 모델을 적용한 방식들보다 평균적으로 4~5% 정도의 성능 개선을 보였다.

I. 서 론

의료 분야에서의 이미지 객체 분할은 진단의 정확성을 높이고 치료 계획을 세우며 질병의 진행을 모니터링 하는데 중요한 역할을 한다. 특히 복부 CT의 경우, 여러 장기가 밀집해 있어 정밀한 분할이 요구되며 이는 병변의 위치, 크기, 그리고 형태를 정확하게 파악할 수 있게 해준다. 이런 객체 분할 분야의 파운데이션 모델(foundation model)인 SAM 모델은 정확하고 효율적인 객체 분할 능력을 입증하였다[1]. 그러나 SAM은 얇은 객체의 분할에 대해서 성능이 떨어지기에, 신뢰성이 높고 요구되는 의료 이미지 객체 분할에서는 보완에 대한 요구가 있어왔다. 본 논문에서는 SAM의 이러한 단점을 보완하기 위해 HQ-SAM 모델을 복부 CT 데이터에 맞게 미세조정(Fine-Tuning)한 모델을 만들었고, 개발된 모델과 기존의 의료 이미지 객체 분할 기법으로 사용되고 있는 SAM, U-Net, DeepLabV3+ 방식의 기존 모델들과 성능 비교 평가를 수행하였다.

II. 본론

1. 복부 CT 이미지 객체 분할 작업에서의 기존 모델

기존 복부 CT 이미지 객체 분할 작업에서는 U-Net 모델과 DeepLabV3+ 모델을 적용해왔다. U-Net 모델은 의료 분야의 이미지 객체 분할을 위해 설계된 아키텍처로, 인코더-디코더 구조를 가지고 있다. 모델의 특징으로는, 인코더와 디코더 사이에 스킵 연결(skip connections)이 있어, 신경망의 초기 층에서 얻은 정보를 다시 사용하여 정확한 위치 정보를 유지하며 객체 분할을 개선했다. DeepLabV3+ 모델은 고급 Atrous Spatial Pyramid Pooling 기법을 사용하며, 다양한 해상도에서 이미지의 세밀한 특성을 포착했다. 이를 통해 의료 이미지에서 정밀한 객체 분할을 실현했다[1]. 그러나 U-Net 모델은 낮은 경계 정확도로 인해 이미지의 세밀한 구조를 식별하는 데 한계가 있었고, DeepLabV3+ 모델은 고해상도 이미지 데이터에서는 해상도를 줄여서 처리해야 하는 문제점이 있었다.

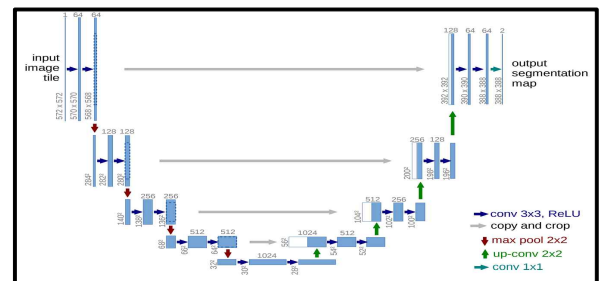


그림 1. U-Net 모델의 구조도

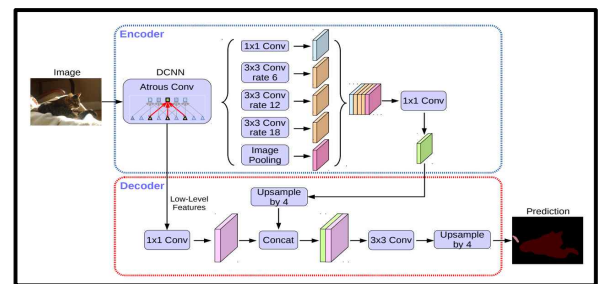


그림 2. DeepLabV3+ 모델의 구조도

2. HQ-SAM의 모델 구조

HQ-SAM 모델은 기존 SAM 모델의 구조를 기반으로 고품질의 객체 분할을 구현했다. SAM 모델은 이미지를 패치로 나누고, MAE(Masked AutoEncoder)의 인코더를 사용하여 이미지 임베딩을 생성하였다. 또한, 프롬프트 인코더를 통해 입력된 프롬프트(점, 박스, 텍스트, 마스크)를 각각의 특성에 맞게 인코딩한다. 마스크 디코더는 임베딩 된 이미지와 프롬프트 정보를 통합하여 최종적인 객체 분할 마스크를 생성했다. SAM은 종종 얇은 객체 구조의 분할을 무시하거나, 깨진 마스크나 까다로운 케이스에 큰 오차가 발생한다는 문제점이 있지만, HQ-SAM은 마스크 디코더에

서 추가로 고해상도 아웃풋 토큰과 글로벌-로컬 퓨전을 도입하여 SAM 모델의 단점을 보완하며, 세밀한 마스크 예측을 가능하게 했다. 이를 통해 이미지의 전체적 구조와 세부적인 디테일을 모두 고려한 고품질의 객체 분할 결과를 제공한다[2].

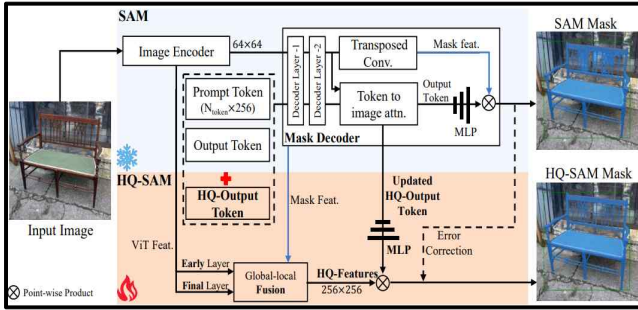


그림 3. HQ-SAM의 모델 구조도

III. 성능 검증 및 실험

1. HQ-SAM 모델 미세 조정 과정

본 논문에서는 사전 학습된 HQ-SAM 모델의 이미지 인코더와 프롬프트 인코더 가중치는 동결(Freezing)하고, 마스크 디코더에 대해 단일 GPU 환경에서 미세 조정을 진행하였다. 학습 데이터셋으로는 MICCAI FLARE22 Challenge Dataset의 80%를 훈련용 데이터로 사용하였고 나머지 20%는 검증용 데이터로 사용하였다. 이 데이터셋은 50개의 복부 CT 스캔으로 구성되어 있으며 각 스캔에는 13개의 장기가 있는 주석 마스크가 포함되어 있다. 손실 함수로는 Dice Loss를 사용하였으며, 이는 Ground Truth 마스크와 Predicted 마스크 사이의 Dice 계수를 기반으로 손실을 계산한다. 옵티마이저는 AdamW를 사용하였다.

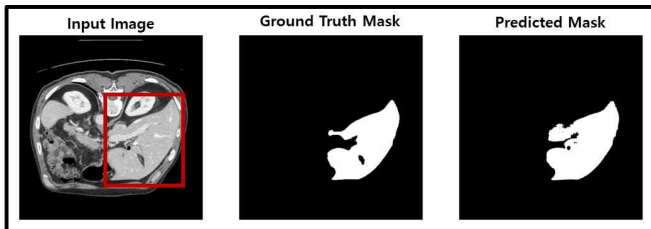


그림 4. 학습 과정 중 Ground Truth 마스크와 모델의 Predicted 마스크

2 성능 검증 결과

본 논문에서는 제안한 방식을 MICCAI FLARE22 Challenge Dataset의 남은 20%에 대해, 평균 Dice Score를 계산함으로써, 모델의 성능을 평가하였다. Dice Score는 일반적인 의료 이미지 객체 분할 성능 측정에 사용되는 평가 지표로 Ground Truth 마스크와 Predicted 마스크 간의 유사도를 측정하는 통계적 방법이다. 이 지표는 0과 1 사이의 값으로 표현되며, 값이 1에 가까울수록 두 마스크 간의 오버랩이 높다는 것을 의미한다. 따라서 Predicted 마스크 결과가 Ground Truth 마스크와 매우 유사할수록 Dice Score가 높아진다. Dice Score는 식 1과 같다[3].

$$Dice = \frac{2 \times |X \cap Y|}{|X| + |Y|} \quad (\text{식 1})$$

식 1에서, X는 Predicted 마스크이며, Y는 Ground Truth 마스크를 의미한다. 본 논문에서 제안한 HQ-SAM 모델을 적용한 방식의 성능을, SAM 모델을 적용한 방식과 기존 의료 이미지 객체 분할 모델의 성능과 비교 평가하기 위하

여, 동일한 데이터셋과 미세 조정을 적용하여 제안한 방식과 SAM, U-NET, DeepLabV3+ 모델을 적용한 방식에 대해 Dice Score를 도출한 후, 비교 평가한 결과는 표 1과 같다.

표 1 각 적용 모델의 검증 결과표

적용모델	Dice Score
SAM	0.9169
U-NET	0.8420
DeepLabV3+	0.8994
제안한 방식(HQ-SAM)	0.9359

표 1의 결과를 보면 제안한 방식의 Dice Score 값이 다른 모델을 적용한 방식들보다 더 높은 값을 나타내는 것을 알 수 있다. 결론적으로, HQ-SAM 모델을 적용한 제안한 방식이 복부 CT 이미지 객체 분할 작업에서 기존의 방식들보다 우수한 성능을 나타낸 것을 확인할 수 있다.

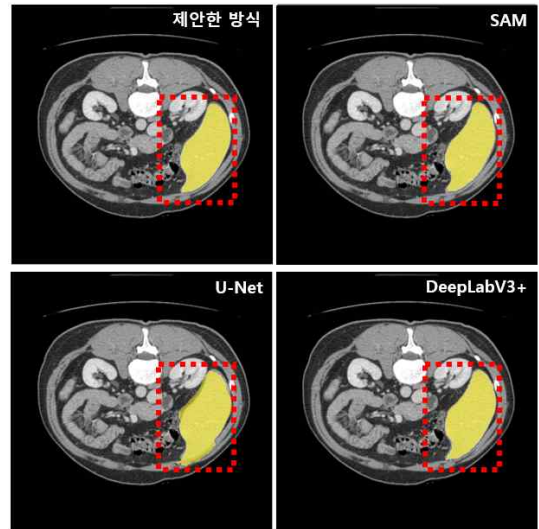


그림 5. 각 적용 모델의 객체 분할 결과

IV. 결론

본 논문에서는 HQ-SAM 모델을 적용하여, 복부 CT 이미지의 객체 분할 작업에서의 성능을 개선시키기 위해 HQ-SAM 모델을 미세 조정하여 도메인에 적합한 모델을 개발하였다. 제안한 모델의 성능을 Dice Score 값을 통해, SAM 모델과 기존의 복부 CT 이미지 객체 분할 모델인 U-Net과 DeepLabV3+ 모델들보다 우수함을 확인할 수 있었다. 본 논문의 결과를 기반으로 본 논문에서 제안한 방식이 복부 질환의 진단과 치료 뿐만 아니라 다양한 의료 이미지에서의 객체 분할 작업의 품질을 개선시킬 수 있는 기술적 기반을 제공할 수 있을 것으로 판단된다. 향후 연구에서는 복부에 한정하지 않고 다양한 부위의 의료 데이터의 HQ-SAM 모델 적용에 관한 연구를 진행하고자 한다.

참 고 문 헌

- [1] Jun Ma et.al., "Segment anything in medical images," Nature Communications 654, 2024
- [2] Lei Ke et.al., "Segment Anything in High Quality," Advances in Neural Information Processing Systems 36 (NeurIPS 2023), 2023
- [3] Fausto Milletari et.al., "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," arXiv:1606.04797, 2016