

3DMedSAM-FDA: 주파수 기반 이중 경로 3D SAM 어댑터

손희경¹, 강윤경², 윤여빈², 정우진³, 차승우⁴

광주과학기술원¹, 성신여자대학교², 홍익대학교³, 서울과학기술대학교⁴

gyoengist@gm.gist.ac.kr, 20232632@sungshin.ac.kr, 20220904@sungshin.ac.kr,
wjjjeong98@g.hongik.ac.kr, tmddn4182@seoultech.ac.kr

3DMedSAM-FDA: Frequency-based Dual-Path Adapter

Hui Gyoeng Son, Yoon Kyeong Kang, Yo Bin Yoon, Woo Jin Jeong, Seung Woo Cha

Gwangju Institute of Science and Technology, Sungshin Women's Univ.,
Hongik Univ., Seoul National University of Science and Technology

요약

본 논문은 3 차원 의료 영상에서의 프롬프트 기반 분할 성능 향상을 목표로 한다. 기존 3D SAM 어댑터[2]의 단일 경로 구조가 국소 정보 및 경계 표현에 한계를 가진다는 점에 차안하여, 전역 문맥 정보와 주파수 기반 국소 정보를 병렬적으로 처리하는 이중 경로 어댑터를 제안한다. 실험 결과, 제안한 방법은 프롬프트 기반 분할에서 기존 대비 일관된 성능 향상을 보였으며, 특히 국소 영역 표현에서 우수한 성능을 확인하였다.

I. 서 론

프롬프트 기반 분할[1]을 3 차원 의료 영상에 적용하려는 연구[2]는 지속적으로 이루어져 왔으며, X-ray 및 CT 와 같은 3 차원 영상은 임상 데이터의 대부분을 차지한다. 그러나 기존 3D SAM 어댑터[2]는 전역적 의미 정보에 비해 국소 구조와 경계 정보를 충분히 반영하지 못한다. 의료 영상에서 종양과 같은 병변은 크기가 작고 경계가 불명확한 경우가 많아 국소 정보의 정밀한 표현이 분할 성능 향상에 필수적이며, 정확한 분할은 절제 범위를 최소화하고, 치료 계획의 정밀화를 통해 환자 예후와 삶의 질을 향상시킨다.

II. 본 론

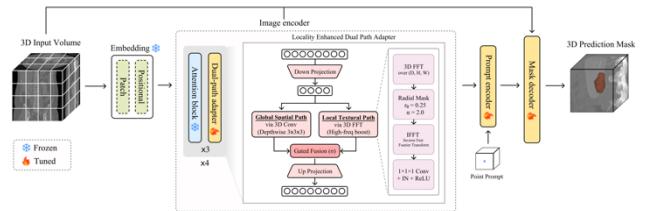
2.1. 배경

3D SAM 어댑터[2]는 2 차원 자연 이미지로 사전 학습된 Segment Anything Model(SAM)[1]을 3 차원 의료 영상에 적용하기 위해 설계된 경량 적용 모듈이다. 전체 구조는 이미지 인코더, 프롬프트 인코더, 마스크 디코더로 구성되며, 이미지 인코더는 전역적 의미 정보를 추출하고 프롬프트 인코더는 사용자 입력을 의미 공간으로 변환한다. 이후 마스크 디코더는 두 정보를 통합해 최종 분할 마스크를 예측한다.

이 3D SAM 어댑터[2]는 기존 SAM의 사전 학습 가중치를 유지한 채 추가적인 어댑터 구조를 통해 3 차원 의료 영상의 도메인 특성을 보정하여 프롬프트 기반 분할을 가능하게 한다. 그러나 단일 경로 기반의 기존 어댑터는 공간 정보의 압축으로 인해 전역 의미 정보에 비해 국소 구조 및 경계 표현에 한계를 가진다.

이러한 한계를 극복하기 위해 영상 처리 분야에서는 주파수 영역 기반 영상 분석에 대한 연구[4]가 증가하고 있다. 일반적으로 저주파 성분은 전반적인 형태와 구조 정보를, 고주파 성분은 경계 및 미세 구조와 같은 세부 정보를 포함하며, 푸리에 변환[4]은 공간 영역 신호를 서로 다른 주파수 성분으로 분해함으로써 이러한 정보를 명시적으로 분리하여 분석할 수 있게 한다. 본 논문에서는 이러한 특성을 활용하여, 주파수 영역에서 고주파 성분을 선택적으로 강조한 뒤 역변환을 통해 공간 영역의 국소 경계 정보를 보강한다.

2.2. 제안 방법



[그림 1] 3DMedSAM-FDA 프레임워크

본 논문은 전역 의미 정보에 편향된 기존 3D SAM 어댑터[2]의 한계를 극복하기 위해, 전역 문맥 경로와 주파수 기반 국소 경로를 병렬로 결합한 이중 경로 어댑터를 제안한다. 두 경로에서 추출된 특징은 전역 경로 기반 계이트를 통해 융합되며, 잔차 연결을 통해 입력 특징을 보정한다.

입력 특징 F 는 채널 축소를 위한 선형 투영을 통해 X 로 변환되며, 이는 위치별 $1 \times 1 \times 1$ 합성곱과 동등하다.

전역 경로는 입력 특징 X 에 대해 $3 \times 3 \times 3$ 깊이별 3 차원 합성곱을 적용하여 3 차원 형태와 장기 문맥 정보를 포착한다. 이를 통해 채널 간 혼합 없이 전역 구조 정보를 포착하며, 기

존 3D 어댑터의 전역적 의미 정렬 특성을 유지한다.

$$F_{global} = \text{ReLU}\left(\text{DWConv}_{3 \times 3 \times 3}(X)\right)$$

국소 경로는 경계 및 미세 구조 정보를 강조하기 위해 주파수 영역 기반 처리를 수행한다. 입력 특징 X 는 3 차원 푸리에 변환을 통해 주파수 도메인 \hat{X} 으로 변환되며, 주파수 좌표의 반경 r 에 기반한 마스크 $M(r)$ 를 적용하여 고주파 성분을 선택적으로 증폭한다. 해당 마스크는 변경 임계값 r_0 이상에서 선형적으로 증가하도록 정의된다.

$$M(r) = 1 + \alpha \cdot \frac{\max(r - r_0, 0)}{1 - r_0}$$

여기서 α 는 고주파 강조 강도를 제어하는 하이퍼파라미터이다. 마스크가 적용된 주파수 특징은 역 푸리에 변환을 통해 다시 공간 영역으로 복원된다.

$$F_{local} = \text{ReLU}\left(\text{IN}\left(\text{Conv}_{1 \times 1 \times 1}\left(F^{-1}(\hat{X} \odot M)\right)\right)\right)$$

이후 $1 \times 1 \times 1$ 합성곱과 정규화 연산을 적용하여, 공간 영역에서 직접적으로 표현하기 어려운 경계 및 미세 구조 정보를 안정적으로 보강한다. 마지막으로 전역 특징으로부터 생성된 계이트를 통해 국소 특징의 기여도를 조절한 후 두 경로의 출력을 융합한다.

$$G = \sigma(F_{global}), \quad F_{fuse} = F_{global} + G \odot F_{local}$$

여기서 $\phi(\cdot)$ 는 $1 \times 1 \times 1$ 합성곱, $\sigma(\cdot)$ 는 시그모이드 함수이다. 최종적으로 선형 복원 연산 $\Psi(\cdot)$ 과 잔차 연결을 적용하여 입력 특징을 다음과 같이 보정한다.

$$F_{out} = X + \Psi(F_{fuse})$$

이러한 이중 경로 구조를 통해 전역적인 형태 및 의미 정렬 능력은 유지하면서도, 주파수 영역에서 강조된 고주파 성분에 기반하여 경계 및 미세 구조 정보가 중요한 영역에서 국소 표현이 선택적으로 강화되도록 한다.

2.3. 데이터셋 및 실험 설정

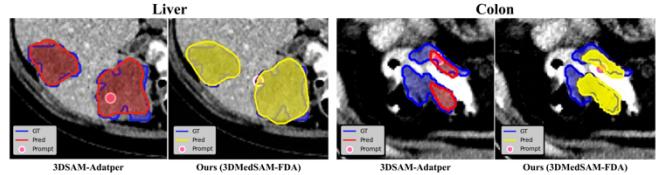
본 연구에서는 KiTS 2021(신장암), MSD-Pancreas(췌장암), LiTS 2017(간암), MSD-Colon(결장암)의 4 가지 3 차원 의료 영상 데이터셋을 사용하였다. 모든 데이터는 동일 해상도로 전처리[2] 후 학습(70%)·검증(10%)·테스트(20%) 세트로 분할하였다. 모델은 AdamW 옵티마이저를 사용하였으며, 학습률은 0.0004로 설정하였다. 학습은 200 에폭 동안 진행되었고, 입력으로는 $128 \times 128 \times 128$ 크기 패치를 무작위 크롭 방식으로 사용하였다. 메모리 제약으로 배치 크기는 1로 설정하였다. 테스트 단계에서는 객체 내부에 제공되는 양성 포인트의 개수에 따라 1, 3, 10 포인트로 성능을 평가하였다. 주파수 기반 국소 경로의 고주파 강조 강도 계수 α 는 2.0, 반경 임계 비율 r_0 은 0.25로 설정하였다.

2.4. 실험 결과 및 분석

Methods	Kidney Tumor		Pancreas Tumor		Liver Tumor		Colon Cancer		#Tuned Params
	Dice ↑	NSD ↑	Dice ↑	NSD ↑	Dice ↑	NSD ↑	Dice ↑	NSD ↑	
nnU-Net (Nat. Methods 2021)	73.07	77.47	41.65	62.54	60.10	75.41	43.91	52.52	30.76M
Swin-UNETR (CVPR 2022)	65.54	72.04	40.57	60.05	50.26	64.32	35.21	42.94	62.19M
3D UX-Net (ICLR 2023)	57.59	58.55	34.83	52.56	45.54	60.67	28.50	32.73	53.01M
<i>1 Point Setting</i>									
SAM-B (1 pt/slice) [1]	36.30	29.86	24.01	26.74	6.71	7.63	28.83	33.63	—
3DSAM-adapter (1 pt/vol) [2]	73.78	83.86	54.09	76.27	54.78	69.55	48.35	63.65	25.46M
Ours (3DMedSAM-FDA, 1 pt/vol)	79.92	86.69	55.58	76.52	56.76	67.98	61.09	76.30	29.02M
<i>3 Points Setting</i>									
SAM-B (3 pts/slice) [1]	39.66	34.85	29.80	33.24	7.87	6.76	35.26	39.31	—
3DSAM-adapter (3 pts/vol) [2]	74.91	84.35	54.92	77.57	56.30	70.02	49.43	65.02	25.46M
Ours (3DMedSAM-FDA, 3 pts/vol)	80.50	88.06	55.40	76.54	60.89	74.33	63.37	78.07	29.02M
<i>10 Points Setting</i>									
SAM-B (10 pts/slice) [1]	40.07	34.96	30.55	32.91	8.56	5.97	39.14	42.70	—
3DSAM-adapter (10 pts/vol) [2]	75.95	84.92	57.47	79.62	56.61	69.52	49.99	65.67	25.46M
Ours (3DMedSAM-FDA, 10 pts/vol)	79.71	87.81	55.81	76.73	61.97	75.71	63.24	78.03	29.02M

[표 1] 실험 결과

실험 결과, 제안한 방법은 장기별 특성에 따라 기존 방법 대비 차별화된 성능 향상을 보였다. 신장암 및 췌장암 분할에서는 기존 3D SAM 어댑터[2] 대비 평균적으로 각 5.1%p, 0.1%p 높은 Dice 분할 성능을 유지하여, 전역 의미 정렬 능력을 저해하지 않으면서 안정적인 결과를 달성하였다.



[그림 2] 3D SAM 어댑터와의 간암, 결장암 분할 결과 비교

간암 분할에서는 포인트 수가 증가함에 따라 성능 향상이 뚜렷하게 나타났으며, 다수의 포인트 설정에서 기존 3D SAM 어댑터[2]뿐만 아니라 nnU-Net[3] 대비 10 포인트 기준 1.87%p 더 높은 Dice 점수를 기록하였다.

결장암 분할에서는 모든 포인트 설정에서 기존 방법 대비 큰 폭의 성능 향상을 달성하여, 기존 모델들이 충분한 성능을 보이지 못하던 결장암 분할 환경에서 제안한 방법의 강점을 가장 명확히 보여준다. 경계가 복잡하고 작은 병변이 빈번한 결장암 분할에서 주파수 기반 국소 경로를 통한 경계 및 미세 구조 보강이 분할 성능 향상에 핵심적으로 기여했음을 확인할 수 있다.

마지막으로, 제안한 방법은 기존 3D SAM 어댑터[2] (25.46M)와 유사한 수준의 파라미터 수(29.02M)를 유지하면서도 이러한 성능 향상을 달성하여, 경량화된 구조 하에서 표현력을 효과적으로 확장할 수 있음을 보여준다.

III. 결 론

본 논문은 3 차원 의료 영상에서 전역 문맥 정보와 국소 특징을 효과적으로 결합하는 3DMedSAM-FDA를 제안하였다. 전역 경로는 해부학적 구조와 의미 정보를 유지하고, 국소 경로는 주파수 영역의 고주파 성분을 강조하여 경계·미세 구조 표현을 강화한다. 실험 결과, 제안한 방법은 프롬프트 기반 분할 환경에서 기존 방법 대비 전반적으로 우수한 성능을 보였으며, 기존 공간 도메인 중심의 3 차원 의료 영상 분할 모델에 주파수 도메인 정보를 효과적으로 적용 가능함을 확인하였다.

참 고 문 헌

- [1] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W. Y., and Dollár, P., "Segment Anything," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4015–4026, 2023.
- [2] Zhang, H., Li, Z., Wang, J., Zhao, Y., and Lu, H., "3D SAM Adapter: Adapting Segment Anything Model for 3D Medical Image Segmentation," arXiv preprint, arXiv:2304.13216, 2023.
- [3] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H., "nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation," Nature Methods, vol. 18, no. 2, pp. 203–211, 2021.
- [4] Yang, Y., Soatto, S., "FDA: Fourier Domain Adaptation for Semantic Segmentation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4085–4095, 2020.