DMQA Open Seminar

Fine-tuning Segment Anything

2025. 01. 03

김성수

Data Mining and Quality Analytics Lab

고려대학교
KOREA UNIVERSITY

발표자 소개

❖ 김성수 (Sungsu Kim)

- 경희대학교 산업경영공학과 학부 졸업 (2022.02)
- 고려대학교 산업경영공학과 대학원 재학
- Data Mining & Quality Analytics Lab. (김성범 교수님)
- 석박통합 과정 (2022.03 ~ Present)

❖ Research Interest

- Computer Vision
- Self/Semi-supervised Learning
- Fine-tuning Foundation Model

❖ Contact

- 2022020650@korea.ac.kr

목차

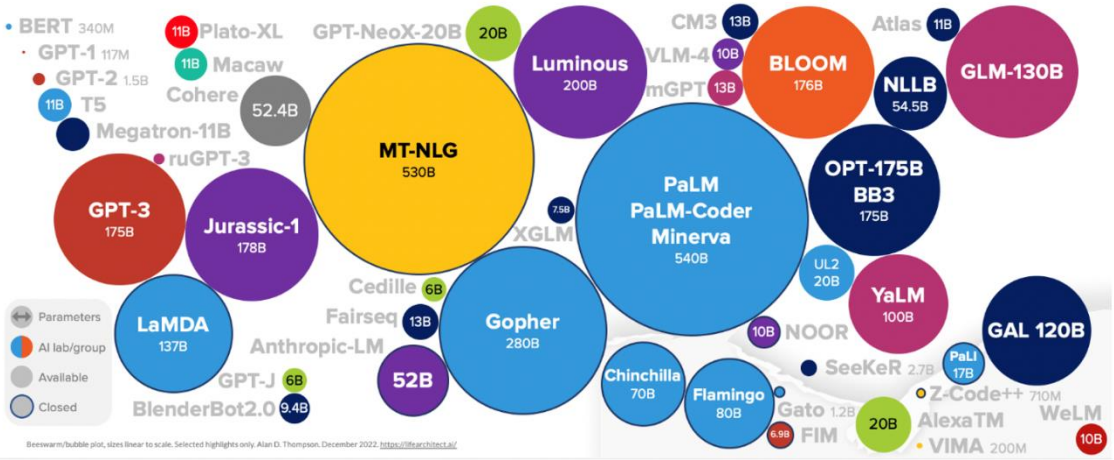
Introduction

Introduction

Foundation 모델

❖ 이제는 Foundation 모델의 시대

- Foundation Model: 다양한 데이터에 대하여 높은 일반화 성능을 갖는 모델
- Transformer 등을 활용한 거대 모델 구조 기반 구조 → 폭넓은 학습 가능
- 수많은(억 단위) 데이터를 활용하여 학습 → 다양한 학습 가능



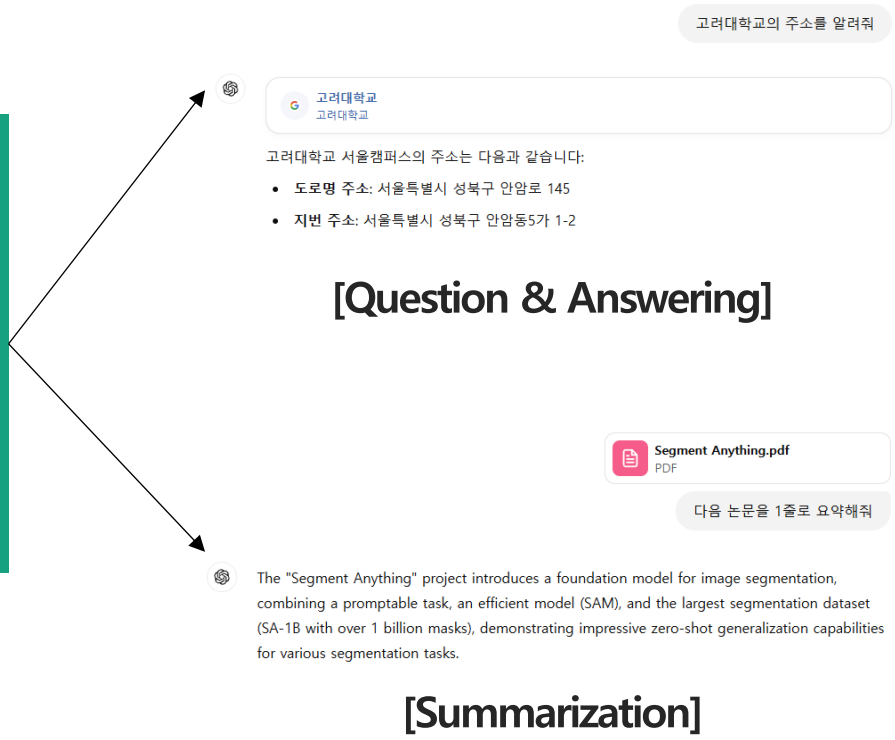
[1] https://petergentsch.com/wp-content/uploads/Language-models-with-the-respective-model-size.png
[2] https://production-media.paperswithcode.com/datasets/ImageNet-0000000008-f2e87edd_Y0fT5zg.jpg

Introduction

Foundation 모델

❖ Foundation 모델 장점

- 추가적인 학습을 하지 않고 다양한 데이터에 적용해도 꽤나 우수한 성능을 보임
- Ex. ChatGPT

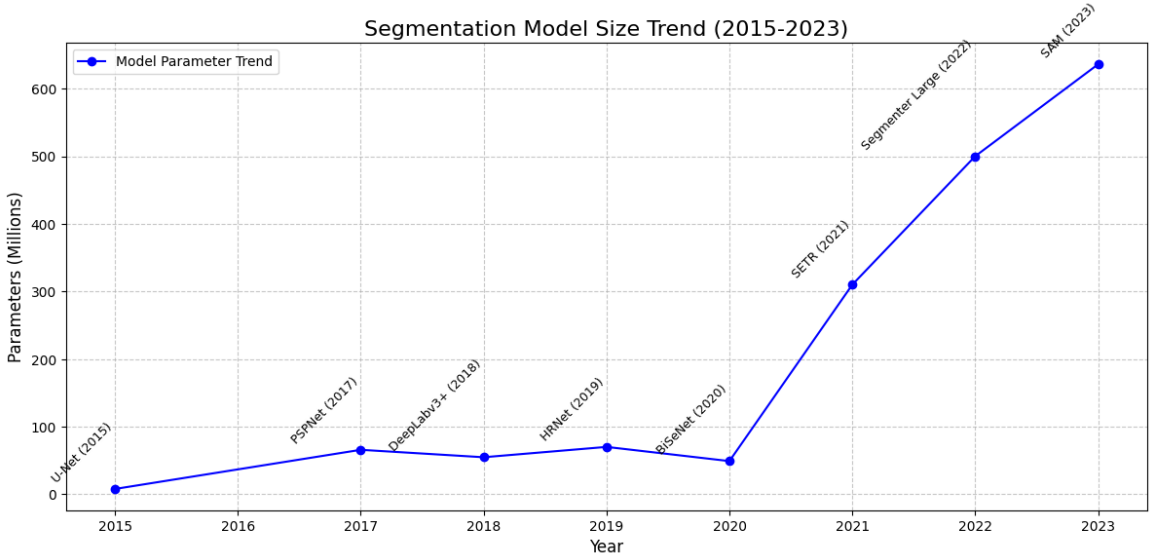


Introduction

Foundation 모델

❖ Foundation 모델 & Image Segmentation

- Image Segmentation 분야에서도 모델 크기는 점점 커지는 추세
- 시간이 지날수록 다양한 데이터셋도 많이 수집 및 레이블링 되고 있음



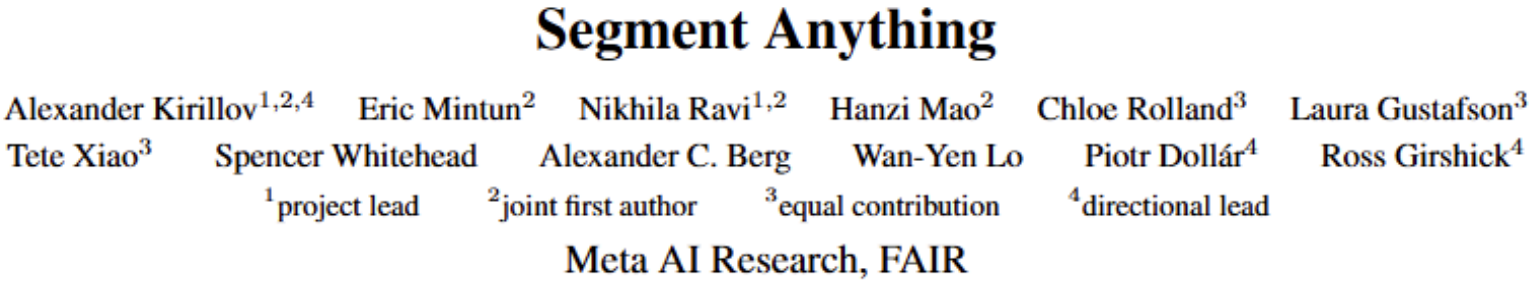
[1] https://eiec.kdi.re.kr/userdata/new_review/14/143/edit/aaa9ZYEUuwrFw4HlxZaKy_1688713400831.png

Introduction

Segmentation Foundation Model: Segment Anything

❖ Segment Anything (SAM)

- Meta에서 2023년 ICCV학회에서 발표 (인용수: 7,613회)
- Segmentation Task에 특화된 Foundation 모델



Segment Anything

Alexander Kirillov[1,2,4] Eric Mintun[2] Nikhila Ravi[1,2] Hanzi Mao[2] Chloe Rolland[3] Laura Gustafson[3]
Tete Xiao[3] Spencer Whitehead Alexander C. Berg Wan-Yen Lo Piotr Dollár[4] Ross Girshick[4]

[1]project lead [2]joint first author [3]equal contribution [4]directional lead

Meta AI Research, FAIR

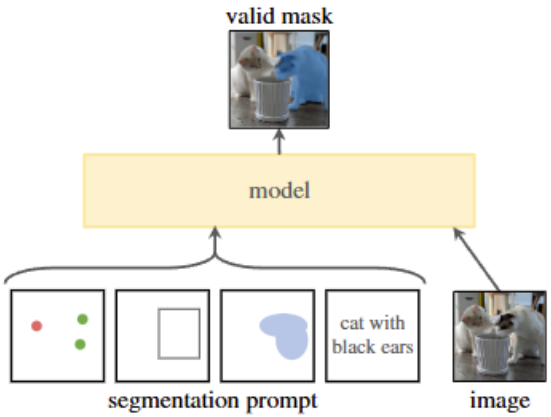
Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4015-4026).

Introduction

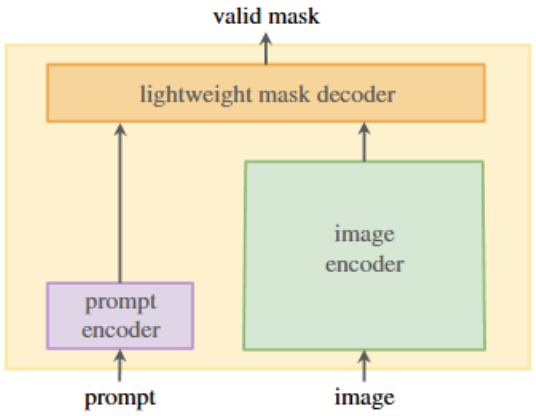
Segmentation Foundation Model: Segment Anything

❖ Segment Anything (SAM)

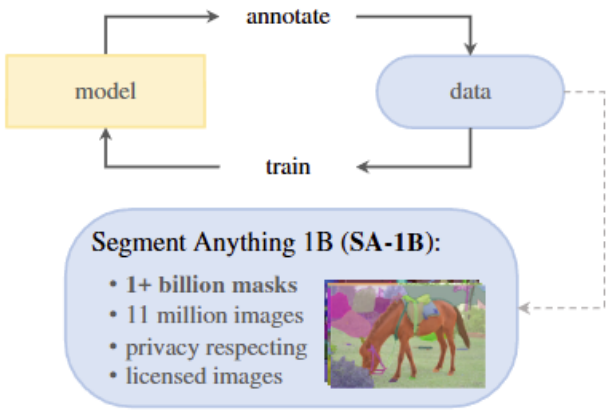
- Interactive Segmentation 모델: 입력 이미지와 사용자의 프롬프트를 함께 입력 받아 예측
- 거대 모델 구조: 6억 3천만개 파라미터를 갖는 Vision Transformer 기반 모델 구조
- 많은 데이터: 웹에서 수집한 이미지를 Semi-auto 방식으로 Labeling한 10억개 Mask로 학습



(a) **Task**: promptable segmentation



(b) **Model**: Segment Anything Model (SAM)



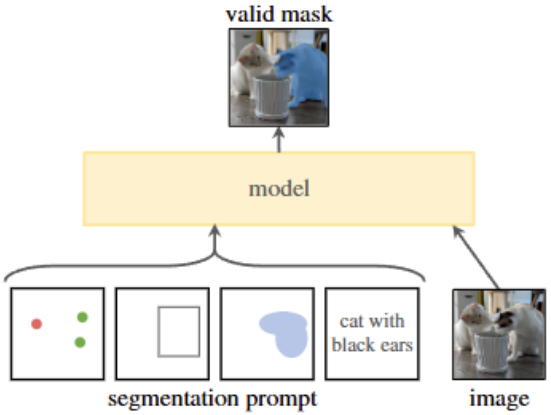
(c) **Data**: data engine (top) & dataset (bottom)

Introduction

Segmentation Foundation Model: Segment Anything

❖ Segment Anything (SAM)

- Interactive Segmentation 모델: 입력 이미지와 사용자의 프롬프트를 함께 입력 받아 예측
- 거대 모델 구조: 6억 3천만개 파라미터를 갖는 Vision Transformer 기반 모델 구조
- 많은 데이터: 웹에서 수집한 이미지를 Semi-auto 방식으로 Labeling한 10억개 Mask로 학습



(a) **Task**: promptable segmentation

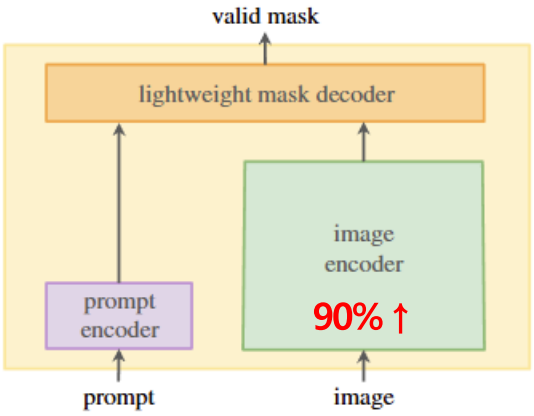
Input Image | Ground Truth | SAM (Point Prompt) | SAM (BBox Prompt)

Introduction

Segmentation Foundation Model: Segment Anything

❖ Segment Anything (SAM)

- Interactive Segmentation 모델: 입력 이미지와 사용자의 프롬프트를 함께 입력 받아 예측
- 거대 모델 구조: 6억 3천만개 파라미터를 갖는 Vision Transformer 기반 모델 구조
- 많은 데이터: 웹에서 수집한 이미지를 Semi-auto 방식으로 Labeling한 10억개 Mask로 학습



(b) **Model**: Segment Anything Model (SAM)

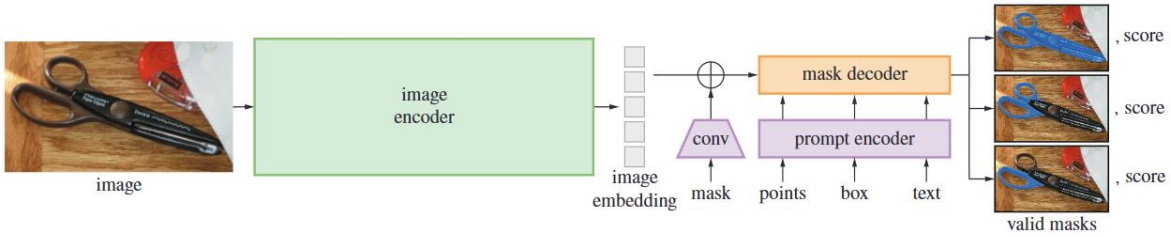


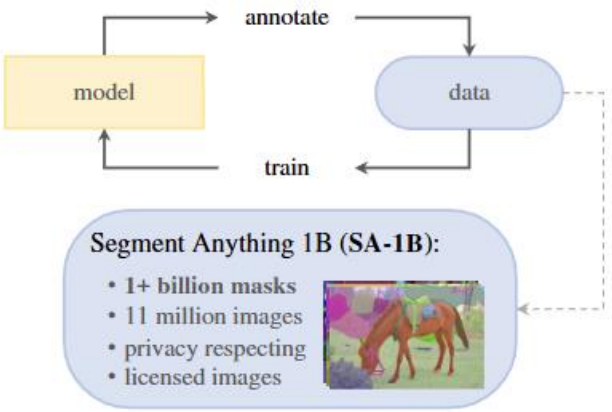
Image Encoder + Prompt Encoder + Mask Decoder로 구성

Introduction

Segmentation Foundation Model: Segment Anything

❖ Segment Anything (SAM)

- Interactive Segmentation 모델: 입력 이미지와 사용자의 프롬프트를 함께 입력 받아 예측
- 거대 모델 구조: 6억 3천만개 파라미터를 갖는 Vision Transformer 기반 모델 구조
- 많은 데이터: 웹에서 수집한 이미지를 Semi-auto 방식으로 Labeling한 약 10억개 Mask로 학습



(c) **Data**: data engine (top) & dataset (bottom)

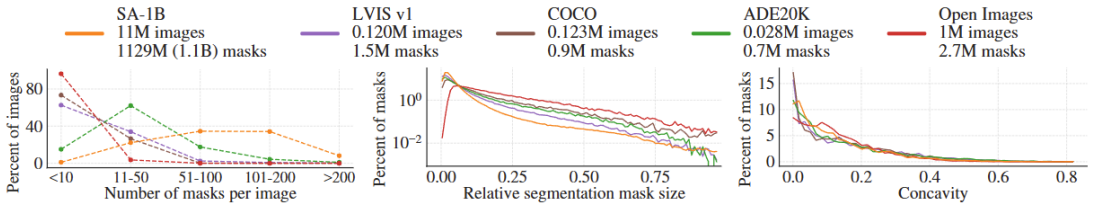


Figure 6: Dataset mask properties. The legend references the number of images and masks in each dataset. Note, that SA-1B has 11× more images and 400× more masks than the largest existing segmentation dataset Open Images [60].

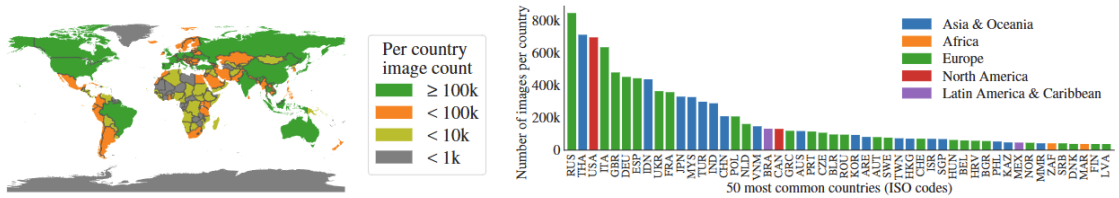
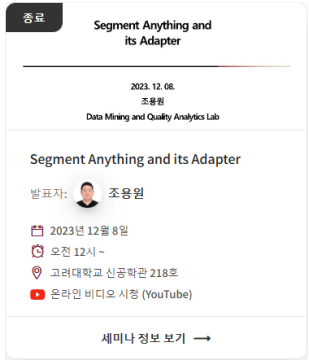


Figure 7: Estimated geographic distribution of SA-1B images. Most of the world's countries have more than 1000 images in SA-1B, and the three countries with the most images are from different parts of the world.

Introduction

Segmentation Foundation Model: Segment Anything



❖ Segment Anything (SAM)

- 객체와 배경의 구분이 뚜렷한 일상 이미지에서 우수한 성능을 보임

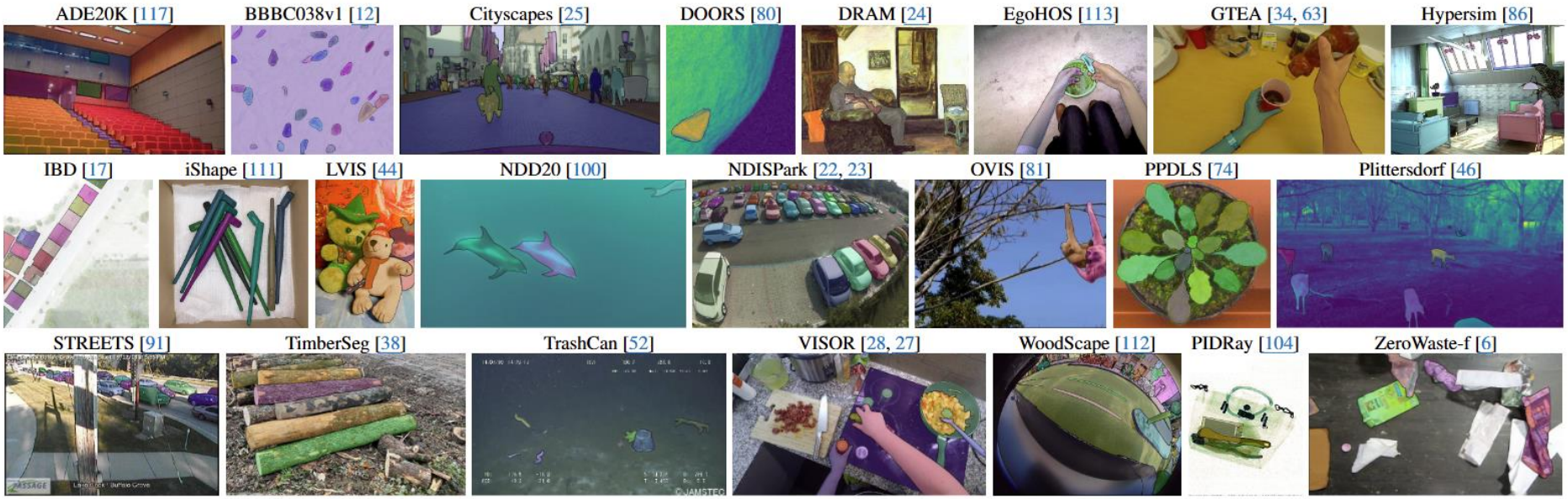


Figure 8: Samples from the 23 diverse segmentation datasets used to evaluate SAM's zero-shot transfer capabilities.

Introduction

Segmentation Foundation Model: Segment Anything

❖ Segment Anything (SAM)

- 객체와 배경의 구분이 뚜렷한 일상 이미지에서 우수한 성능을 보임

[Question]

그렇다면, SAM은 모든 데이터에서 좋은 성능을 보일 것인가?

Introduction

SAM의 한계

❖ **SAM은 모든 것을 Segmentation 할 수 있는 것은 아니다.**

- 2024년 Machine Intelligence Research저널에서 발표 (인용수: 177회)
- 객체가 뚜렷하게 안 보이거나, 일상적인 이미지가 아닌 이미지에 SAM은 저조함

Segment Anything Is Not Always Perfect: An Investigation of SAM on Different Real-world Applications

Wei Ji[1] Jingjing Li[1] Qi Bi[2] Tingwei Liu[3] Wenbo Li[4] Li Cheng[1]

[1]University of Alberta, Edmonton T6G 2R3, Canada

[2]Wuhan University, Wuhan 430072, China

[3]Dalian University of Technology, Dalian 116024, China

[4]Samsung Research America, Mountain View 94043, USA

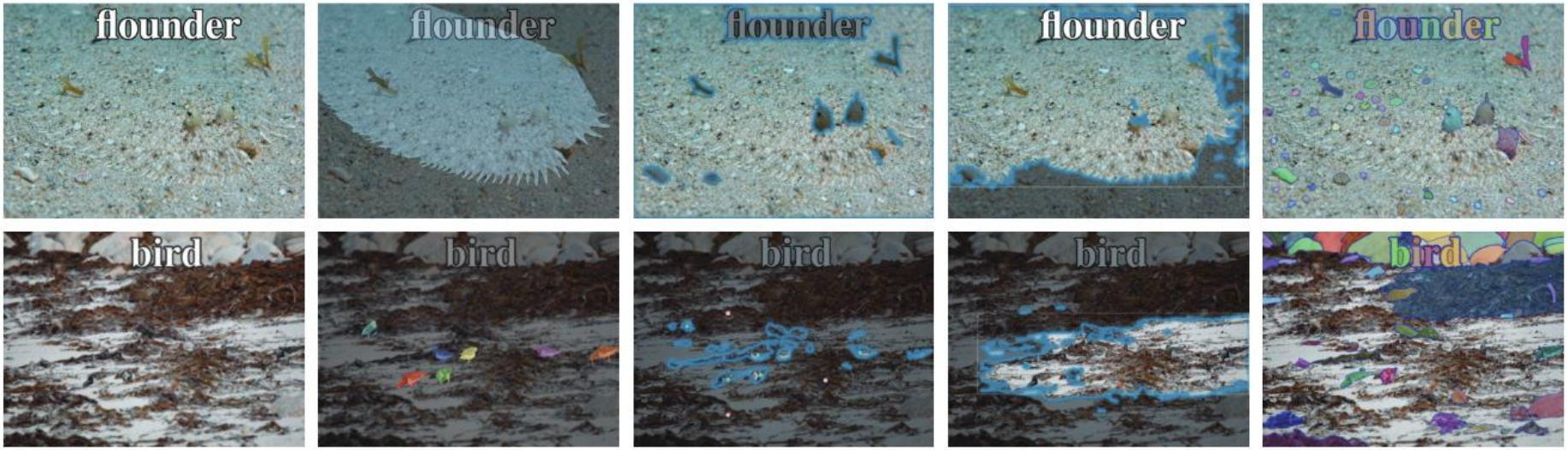
Ji, W., Li, J., Bi, Q., Liu, T., Li, W., & Cheng, L. (2024). Segment anything is not always perfect: An investigation of sam on different real-world applications. Machine Intelligence Research.

Introduction

SAM의 한계

❖ Failure Case① 배경과 객체가 뚜렷하지 않은 Case

- 배경과 뚜렷하게 구분되는 객체에**"만"** 우수
- 배경과 객체가 뚜렷하게 구분되지 않는 경우, 저조한 성능을 보임
  - ➢ (개인의견) SAM의 Semi-auto에서, 이러한 이미지들은 불완전한 레이블을 가질 것 → 우수한 학습이 어려움



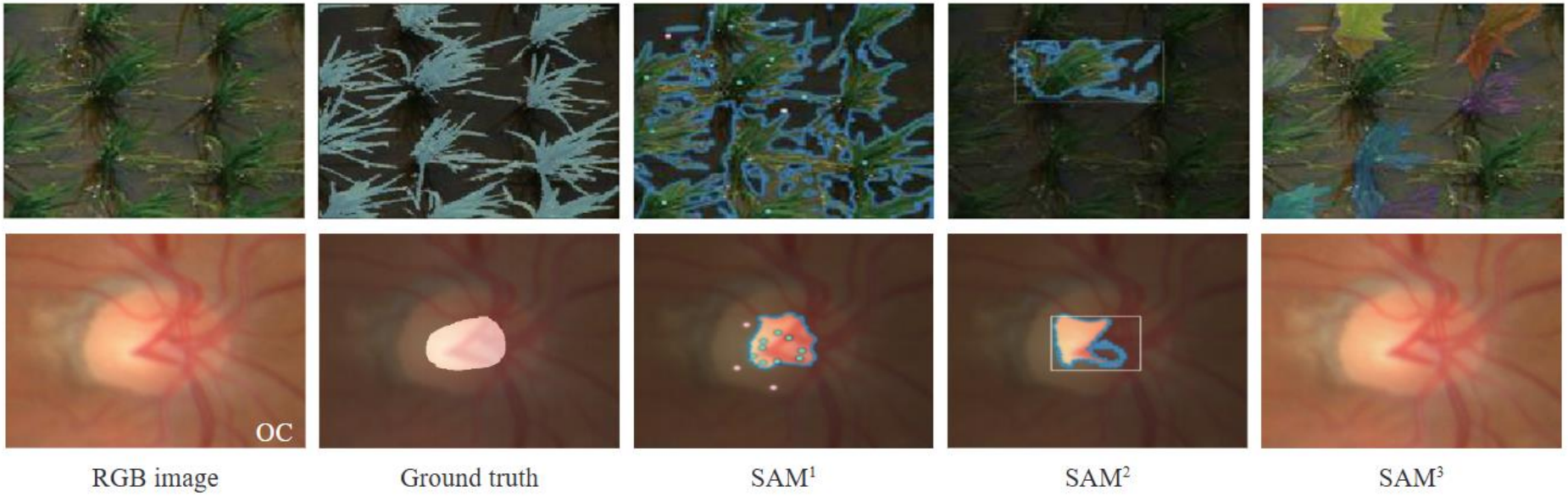
RGB image | Ground truth | SAM[1] | SAM[2] | SAM[3]

Introduction

SAM의 한계

❖ Failure Case② Domain-specific 데이터

- 수집 및 학습이 어려운 의료, 농업 등 Domain-specific 데이터에서 저조한 성능을 보임
  - ➢ 농업 데이터: 작물이나, 토지들은 주로 객체가 아닌 배경으로 레이블링 되기에, 탐지가 어려움
  - ➢ 의료 데이터: 의료 데이터는 수집에 큰 비용이 필요하기에, 충분한 학습이 되지 못함



Introduction

SAM의 한계

❖ Failure Case② Domain-specific 데이터

- 수집 및 학습이 어려운 의료, 농업 등 Domain-specific 데이터에서 저조한 성능을 보임
  - ➢ 의료 데이터: 의료 데이터는 수집에 큰 비용이 필요하기에, 충분한 학습이 되지 못함
  - ➢ 농업 데이터: 작물이나, 토지들은 주로 객체가 아닌 배경으로 레이블링 되기에, 탐지가 어려움

[Question]

어떻게 SAM을 개선할 수 있을까?

Introduction

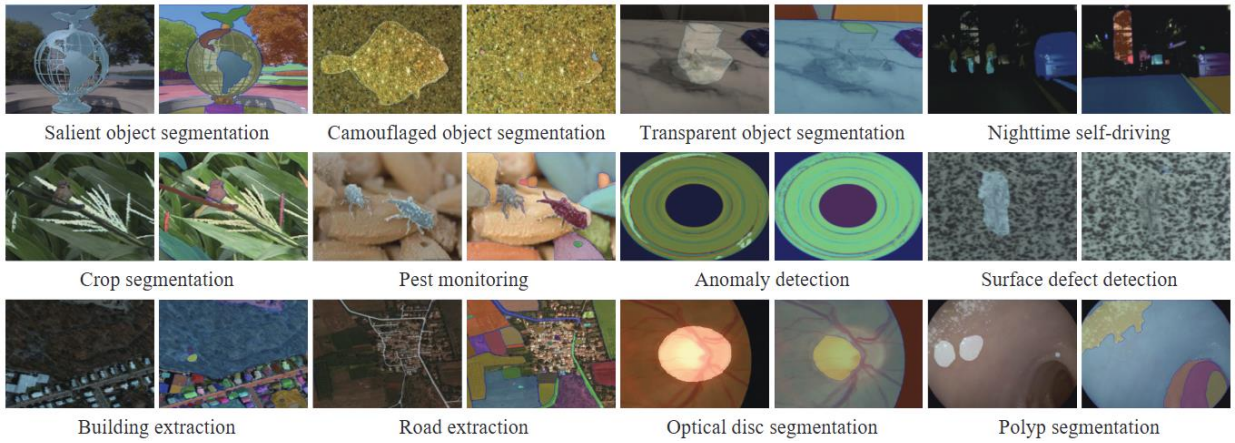
How to Improve SAM?

❖ SAM Fine-tuning을 통한 성능 개선

- 갖고 있는 데이터를 기반으로, SAM을 추가적으로 학습

Pretrained SAM

자신에게 주어진 Task 데이터로 SAM을 추가 학습



① SAM의 Upper Bound 개선 가능

② 정교한 Prompt 불필요

Algorithms

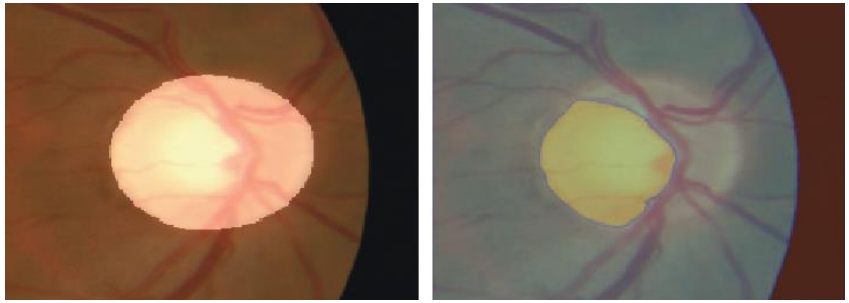
Algorithms

들어가기 전에

❖ 핵심: SAM을 어떻게 Fine-tuning했는지에 집중해서 청취할 것

- "어떻게 SAM을 Fine-tuning 했는지"에 초점을 맞추어 청취할 것을 권장
- 각 Domain에 특화된 아이디어를 제안하지는 않음
  - ➢ Benchmark에서는 모두 잘 나오기에, Domain-specific 데이터를 활용한 것 → 해당 도메인 문제를 해결하려는 목적은 아님

농작물 Segmentation



의료 Segmentation

Algorithms

(1) Segment Anything in Medical Images (MedSAM)

❖ Segment Anything in Medical Images (MedSAM)

- 2023년 Nature Communications저널에서 발표 (인용수: 1,151회)
- SAM을 의료 이미지로 추가적인 학습 수행 → 성능 개선

nature communications

Article

https://doi.org/10.1038/s41467-024-44824-z

Segment anything in medical images

Received: 24 October 2023

Accepted: 5 January 2024

Published online: 22 January 2024

Check for updates

Jun Ma[1,2,3], Yuting He[4], Feifei Li[1], Lin Han[5], Chenyu You[6] & Bo Wang[1,2,3,7,8]

Medical image segmentation is a critical component in clinical practice, facilitating accurate diagnosis, treatment planning, and disease monitoring.

Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment anything in medical images. Nature Communications, 15(1), 654.

Algorithms

(1) Segment Anything in Medical Images (MedSAM)

❖ MedSAM 배경

- **Domain-specific 이미지에 대해 충분히 학습되지 못한 SAM은 의료 이미지에서 저조**
- 해결방향: 의료 이미지를 활용하여 SAM을 Fine-tuning 하자!

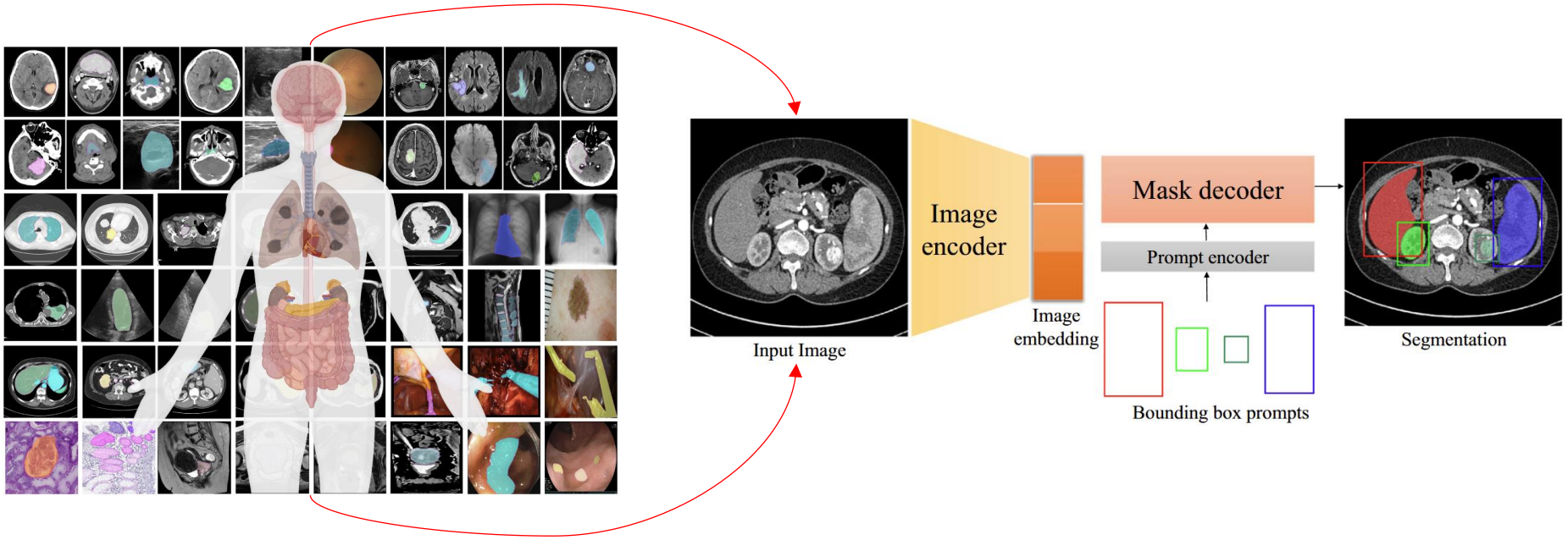
[1] https://analyticsindiamag.com/wp-content/uploads/2021/03/pasted-image-0-11.png
[2] https://production-media.paperswithcode.com/datasets/1b0ca5c4-4b61-4cf7-9d9c-48a2aa5bba3f.png
[3] https://production-media.paperswithcode.com/thumbnails/task/task-0000000876-6fbe75a2_gBlYteG.jpg
[4] https://www.researchgate.net/publication/339873650/figure/fig1/AS:868106524168192@1583984132224/Different-applications-of-medical-image-segmentation.png

Algorithms

(1) Segment Anything in Medical Images (MedSAM)

❖ MedSAM 방법론

- 의료 이미지 수집: 다양한 해부학적 구조를 포함하는 의료 이미지 150만장 수집
- 모델 학습: SAM 모델 전체를 Fine-tuning

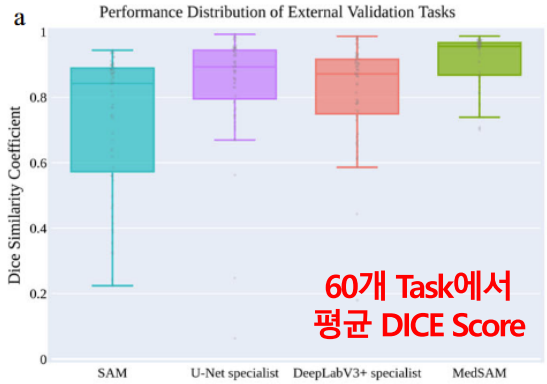


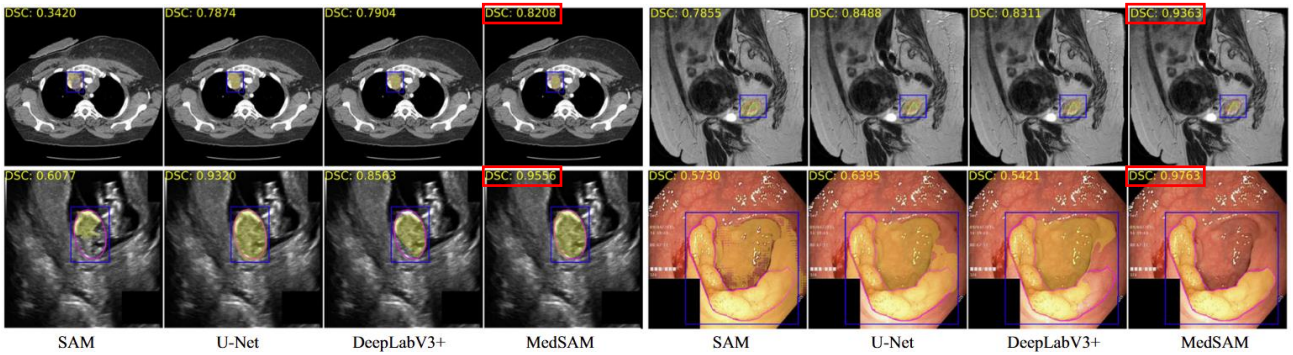
Algorithms

(1) Segment Anything in Medical Images (MedSAM)

❖ MedSAM 실험결과

- 학습하지 않았던 60종류 의료 이미지에서 우수한 성능 및 가장 작은 편차를 보임
- 시각적으로도 우수한 성능을 보임



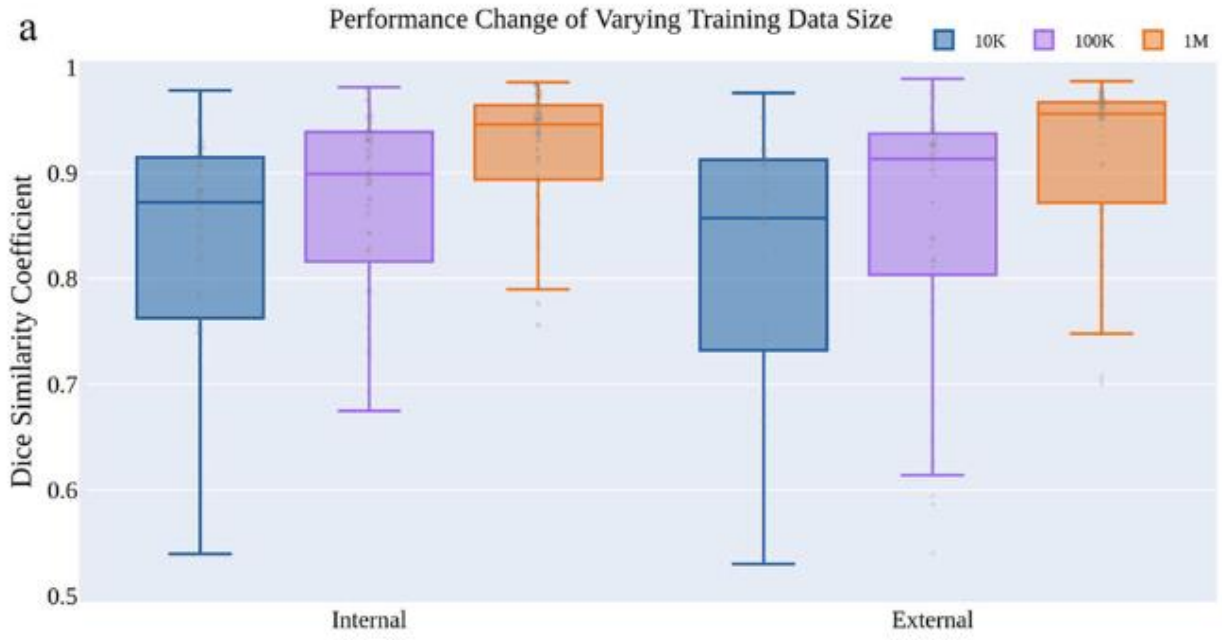


Algorithms

(1) Segment Anything in Medical Images (MedSAM)

❖ MedSAM 실험결과

- 학습 데이터가 많아질수록 비교적 안정적인 성능 확보 가능
  - ➢ Vanilla SAM의 Dice Score는 0.2 → 10K 데이터만으로도 성능 개선 가능



Algorithms

(1) Segment Anything in Medical Images (MedSAM)

❖ MedSAM 한계

- SAM 전체를 Fine-tuning 하는 것은 많은 GPU가 필요 (SAM Original Paper 기준 A100 256장)
- 따라서, 일반적인 상황에서 MedSAM으로 모델을 학습하는 것은 현실적으로 거의 불가능

Ethical Considerations

Data	We trained SAM on licensed images. The images were filtered for objectionable content by the provider, but we acknowledge the possibility of false negatives. We performed a geographic analysis of the SA-1B dataset in §6. While SA-1B is more geographically diverse than many of its predecessors, we acknowledge that some geographic regions and economic groups are underrepresented.
Cost and impact of compute	SAM was trained on 256 A100 GPUS for 68 hours. We acknowledge the environmental impact and cost of training large scale models. The environmental impact of training the released SAM model is approximately 6963 kWh resulting in an estimated 2.8 metric tons of carbon dioxide given the specific data center used, using the calculation described in [77] and the ML CO_2 Impact calculator [61]. This is equivalent to ~7k miles driven by the average gasoline-powered passenger vehicle in the US [101]. We released the SAM models to both reduce the need for retraining and lower the barrier to entry for large scale vision research.
Risks and harms	We evaluated SAM for fairness in §6. Downstream use cases of SAM will create their own potential for biases and fairness concerns. As such we recommend users run their own fairness evaluation when using SAM for their specific use case.
Use cases	We implore users to use their best judgement for downstream use of the model.

Algorithms

(1) Segment Anything in Medical Images (MedSAM)

❖ MedSAM 한계

- SAM 전체를 Fine-tuning 하는 것은 많은 GPU가 필요 (SAM Original Paper 기준 A100 256장)
- 따라서, 일반적인 상황에서 MedSAM으로 모델을 학습하는 것은 현실적으로 거의 불가능

[Question]

효율적으로 SAM을 Fine-tuning 할 수 없을까?

Algorithms

(2) Sam-Adapter: Adapting Segment Anything in Underperformed Scenes

❖ Sam-Adapter: Adapting Segment Anything in Underperformed Scenes

- 2023년 ICCV에서 발표 (인용수: 277회(단, arXiv + ICCV 버전 모두 포함))
- Domain-specific 데이터에 SAM을 **효율적으로** Fine-tuning

SAM-Adapter: Adapting Segment Anything in Underperformed Scenes

Tianrun Chen[1, 2, +] Lanyun Zhu[4, +] Chaotao Ding[3, +] Runlong Cao[3 +] Yan Wang[5]
Shangzhan Zhang[1] Zejian Li [1] Lingyun Sun [1] Ying Zang [3, *] Papa Mao [2]

Zhejiang University[1] KOKONI, Moxin (Huzhou) Technology[2]
Huzhou University[3] Singapore University of Technology and Design[4] Beihang University [5]
{tianrun.chen, zhang3z, zejianlee, sunly}@zju.edu.cn lanyun_zhu@mymail.sutd.edu.sg
2021388117@stu.zjhu.edu.cn crl1657@163.com
wangyan9509@gmail.com info@kokoni3d.com 02750@zjhu.edu.cn

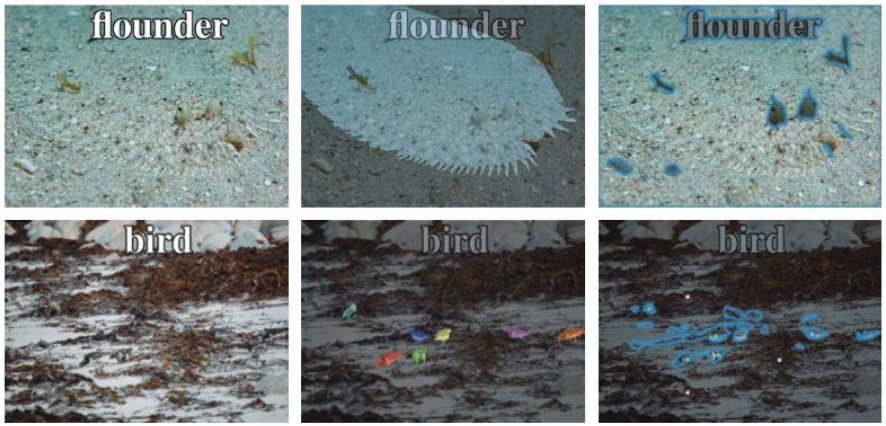
Chen, T., Zhu, L., Deng, C., Cao, R., Wang, Y., Zhang, S., ... & Mao, P. (2023). Sam-adapter: Adapting segment anything in underperformed scenes. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3367-3375).

Algorithms

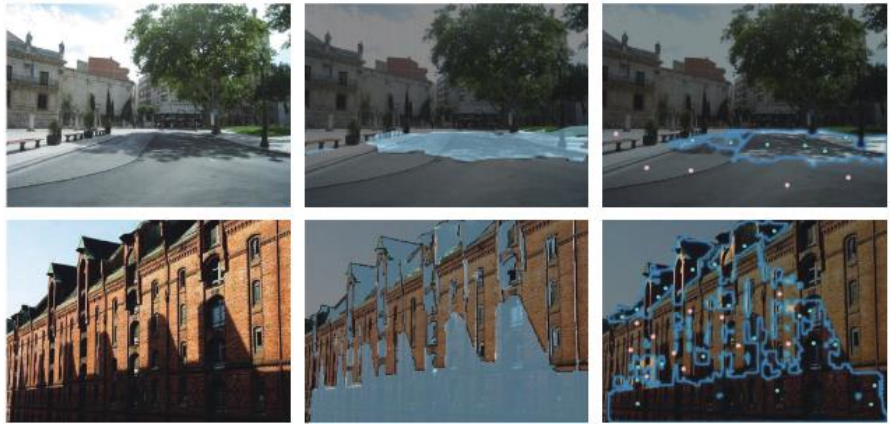
(2) Sam-Adapter: Adapting Segment Anything in Underperformed Scenes

❖ Sam-Adapter 배경

- SAM은 특수한 Segmentation Task에서는 저조한 성능을 보임
- **전체 모델을 Fine-tuning 하는 것은 큰 Resource가 필요**



Camouflaged Detection



Shadow Detection

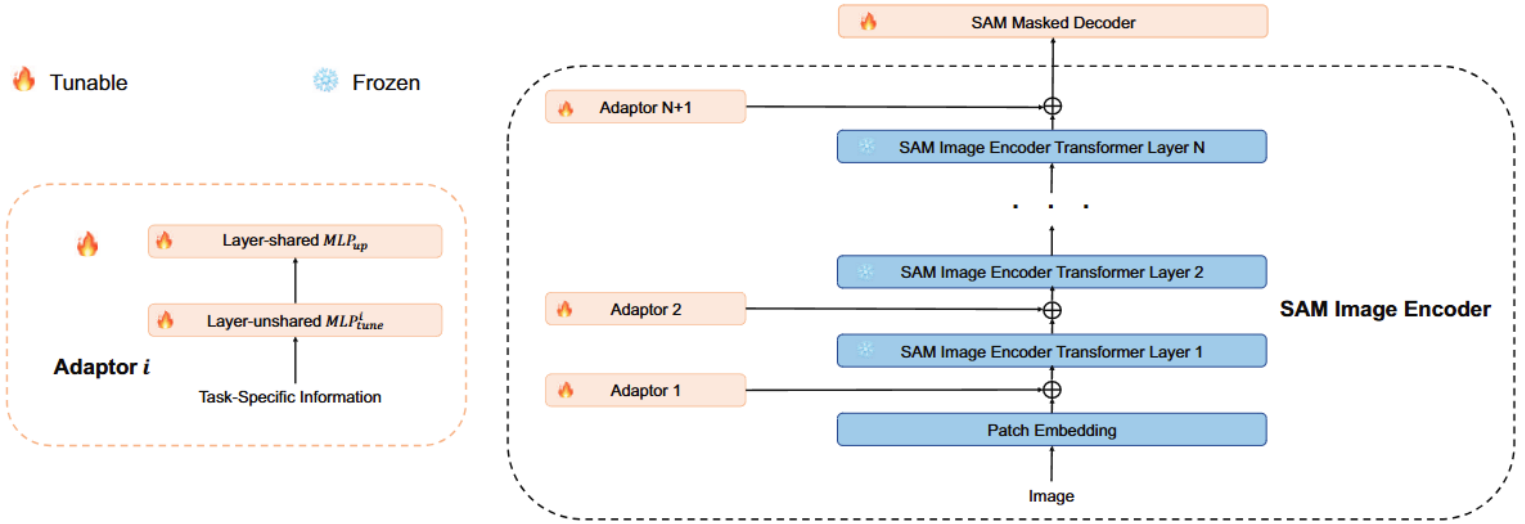
어떻게 이러한 특수 Domain에 Foundation 모델을 효율적으로 활용할 수 있을까?

Algorithms

(2) Sam-Adapter: Adapting Segment Anything in Underperformed Scenes

❖ Sam-Adapter 방법론

- Motivation: SAM이 갖고 있는 훌륭한 지식에, **외부 지식을 주입**해보자. (SAM 대비 5% Param만 학습)
  - SAM Encoder: 기존 SAM Weight는 Freeze, Encoder에 추가된 Adapter Parameter만 Train
  - SAM Decoder: Train
  - Prompt Encoder: 사용X (SAM-adapter는 프롬프트를 활용하지 않는 Segmentation 모델로 활용)

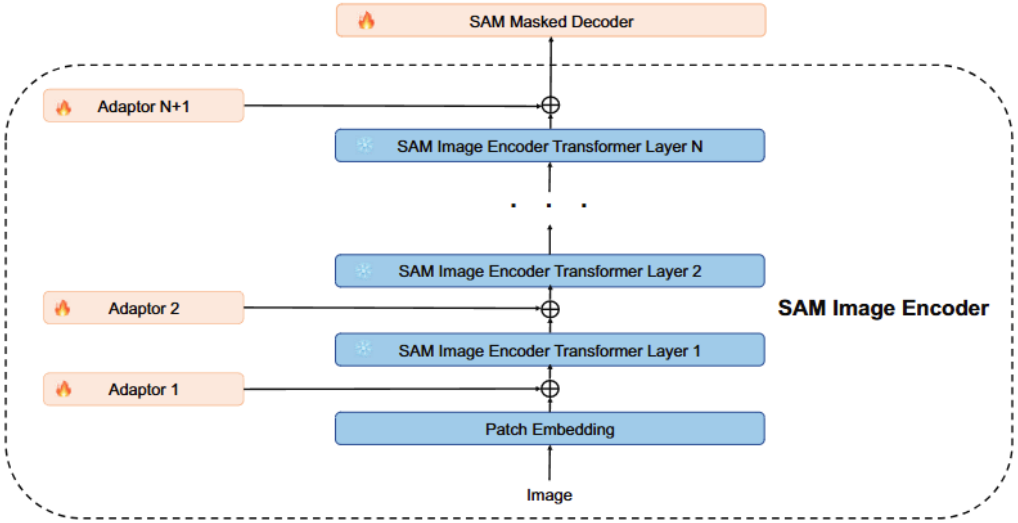


Algorithms

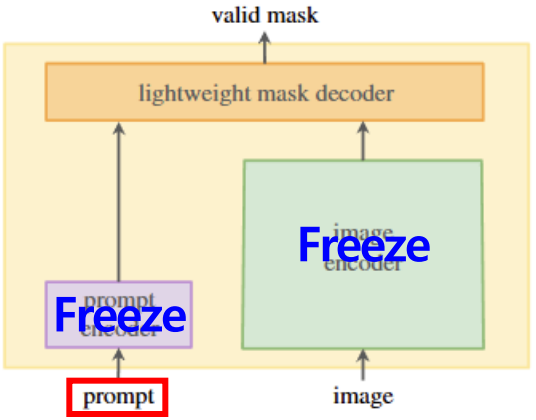
(2) Sam-Adapter: Adapting Segment Anything in Underperformed Scenes

❖ Sam-Adapter 방법론

- Motivation: SAM이 갖고 있는 훌륭한 지식에, **외부 지식을 주입**해보자. (SAM 대비 5% Param만 학습)
  - ➢ SAM Encoder: 기존 SAM Weight는 Freeze, Encoder에 추가된 Adapter Parameter만 Train
  - ➢ SAM Decoder: Train
  - ➢ Prompt Encoder: 사용X (SAM-adapter는 프롬프트를 활용하지 않는 Segmentation 모델로 활용)



SAM-Adapter



모든 이미지에 동일한 프롬프트 삽입 → [0, 0, … 0]

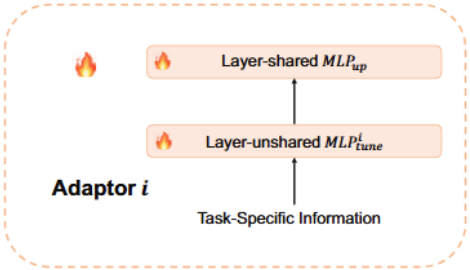
SAM

Algorithms

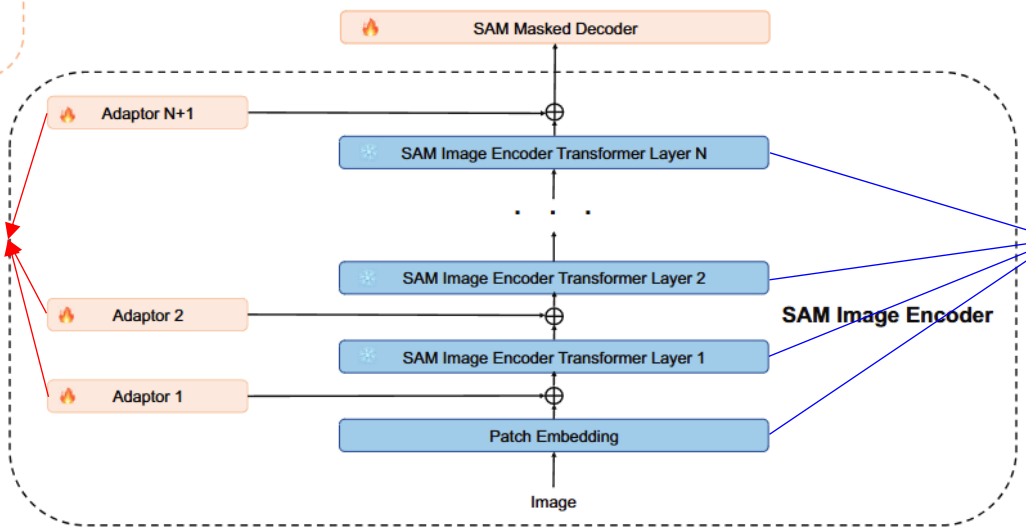
(2) Sam-Adapter: Adapting Segment Anything in Underperformed Scenes

❖ Sam-Adapter 방법론

- SAM Encoder: SAM이 **갖고 있는 기존 지식을 유지**하면서, **외부 지식을 함께 학습**
- **외부지식 학습**에는 MLP 기반 Adapter를 활용
  - ➢ MLP_{tune}: **외부지식을 학습**하는 Layer로, 32개 MLP로 구성
  - ➢ MLP_{up}: Embedding 차원을 맞춰주기 위한 Layer로, 1개 MLP로 구성 (Layer 단위로 Share)



$$P^i = \mathrm{MLP}_{up}\left(\mathrm{GELU}\left(\mathrm{MLP}^i_{tune}\left(F_i\right)\right)\right)$$



[외부 지식]
새로운 데이터를 기반으로 주입 (Train)

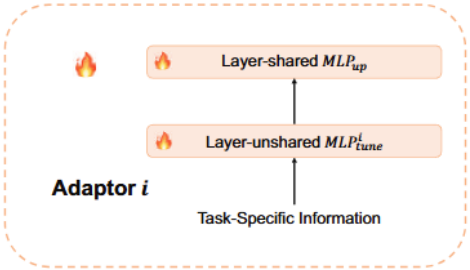
[SAM이 갖고 있던 지식]
기존에 학습된 것 유지 (Freeze)

Algorithms

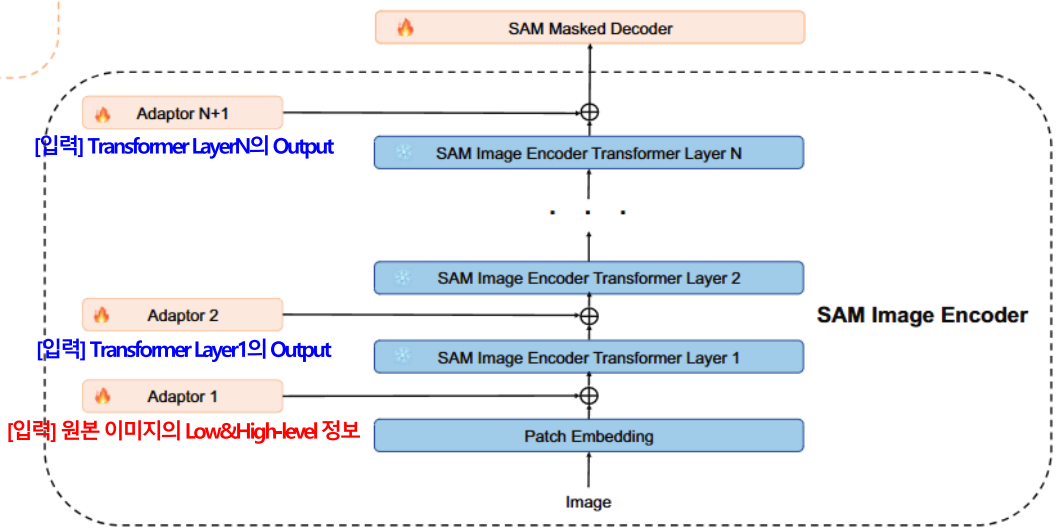
(2) Sam-Adapter: Adapting Segment Anything in Underperformed Scenes

❖ Sam-Adapter 방법론

- **Adapter1**에는 **Task Specific Information** 입력, 그 외 Adapter는 이전 Transformer Output 입력
- Task Specific Information은 Low-level 정보와 High-level 정보를 모두 활용 → Task 정보 극대화
  - ➢ Low-level 정보: Patch Embedding (Linear) + Positional Encoding
  - ➢ High-level 정보: 주어진 이미지를 Fourier Transformation (FFT)한 정보



$$Task\ specific\ Info = f_{low} + f_{high}$$
$$f_{low} = Linear(x) + Positional\ Encoding$$
$$f_{high} = FFT(x)$$



Algorithms

(2) Sam-Adapter: Adapting Segment Anything in Underperformed Scenes

❖ Sam-Adapter 실험결과

- Camouflage Detection 및 Shadow Detection에서 모두 우수한 성능을 보임
  - ➢ SAM은 전체 이미지 크기와 동일한 Box 프롬프트를 넣어서 산출
  - ➢ SAM-Adapter는 별도 프롬프트를 활용하지 않음

Method	CHAMELEON [43]				CAMO [27]				COD10K [12]			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	MAE ↓	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	MAE ↓	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	MAE ↓
SINet[13]	0.869	0.891	0.740	0.440	0.751	0.771	0.606	0.100	0.771	0.806	0.551	0.051
RankNet[36]	0.846	0.913	0.767	0.045	0.712	0.791	0.583	0.104	0.767	0.861	0.611	0.045
JCOD [28]	0.870	0.924	-	0.039	0.792	0.839	-	0.82	0.800	0.872	-	0.041
PFNet [38]	0.882	**0.942**	0.810	0.330	0.782	0.852	0.695	0.085	0.800	0.868	0.660	0.040
FBNet [32]	0.888	0.939	**0.828**	**0.032**	0.783	0.839	0.702	0.081	0.809	0.889	0.684	0.035
SAM [24]	0.727	0.734	0.639	0.081	0.684	0.687	0.606	0.132	0.783	0.798	0.701	0.050
SAM-Adapter (Ours)	**0.896**	0.919	0.824	0.033	**0.847**	**0.873**	**0.765**	**0.070**	**0.883**	**0.918**	**0.801**	**0.025**

Table 1. Quantitative Result for Camouflage Detection

Method	BER ↓
Stacked CNN [46]	8.60
BDRAR [59]	2.69
DSC [20]	3.42
DSD [53]	2.17
FDRNet [61]	1.55
SAM [24]	40.51
SAM-Adapter (Ours)	**1.43**

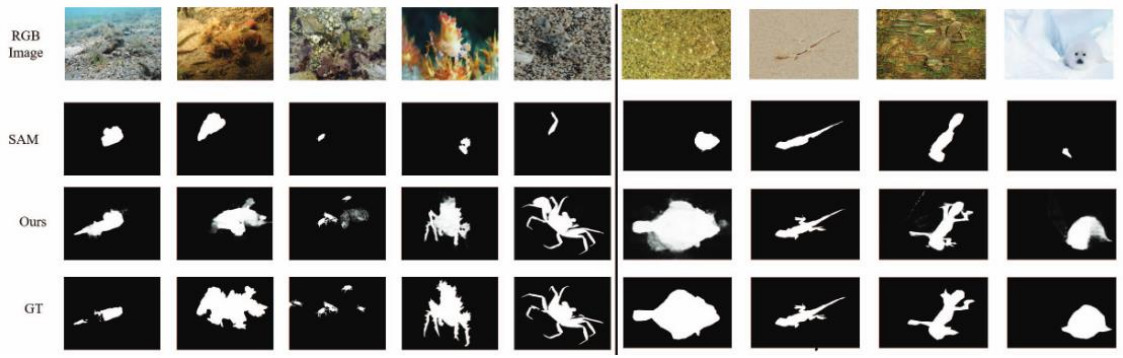
Table 2. Quantitative Result - Shadow Detection

Algorithms

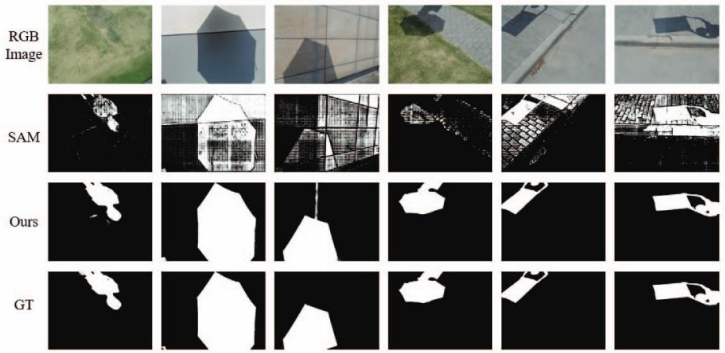
(2) Sam-Adapter: Adapting Segment Anything in Underperformed Scenes

❖ Sam-Adapter 실험결과

- Camouflage Detection 및 Shadow Detection에서 모두 우수한 성능을 보임



[Camouflage Detection]



[Shadow Detection]

Algorithms

(2) Sam-Adapter: Adapting Segment Anything in Underperformed Scenes

❖ Sam-Adapter Ablation Study

- 단순 Decoder만 튜닝하는 것보다, **Encoder 내 Adapter를 함께 튜닝 시 성능이 크게 개선**

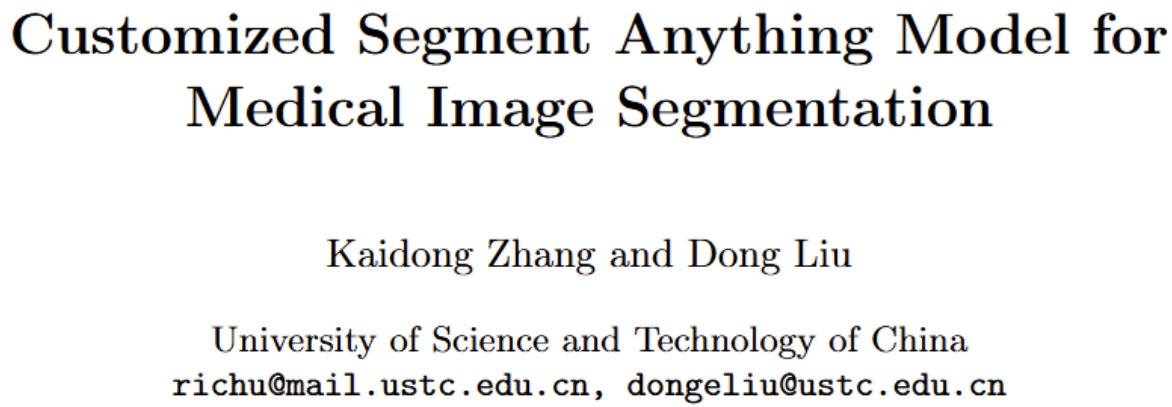
Method	CHAMELEON [43]				CAMO [27]				COD10K [12]			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	MAE $\downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	MAE $\downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	MAE $\downarrow$
w/o SAM-Adapter	0.796	0.802	0.676	0.062	0.750	0.756	0.639	0.105	0.789	0.817	0.596	0.049
w/ SAM-Adapter	**0.896**	**0.919**	**0.824**	**0.033**	**0.847**	**0.873**	**0.765**	**0.070**	**0.883**	**0.918**	**0.801**	**0.025**

Algorithms

(3) Customized Segment Anything Model for Medical Image Segmentation (SAMed)

❖ Customized Segment Anything Model for Medical Image Segmentation (SAMed)

- 2023년 arXiv에 수록 (인용수: 301회)
- SAM을 **SAM-Adapter보다 효율적으로** Fine-tuning



Customized Segment Anything Model for Medical Image Segmentation

Kaidong Zhang and Dong Liu

University of Science and Technology of China
richu@mail.ustc.edu.cn, dongeliu@ustc.edu.cn

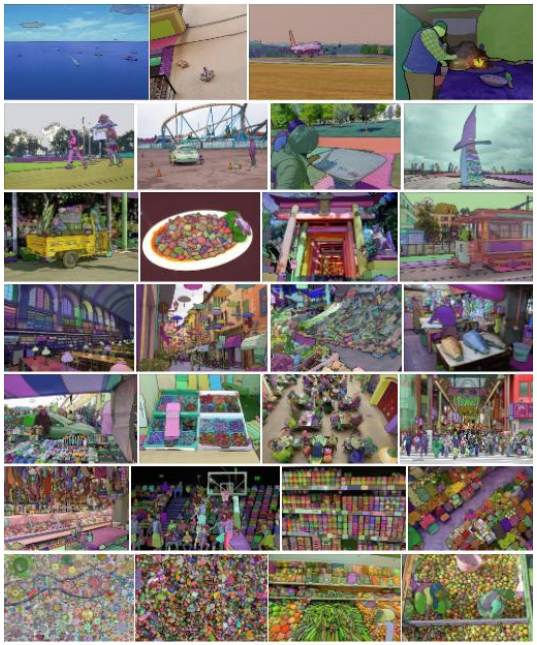
Zhang, K., & Liu, D. (2023). Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785.

Algorithms

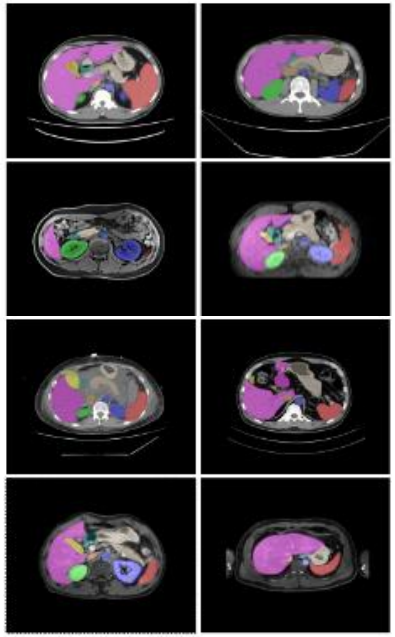
(3) Customized Segment Anything Model for Medical Image Segmentation (SAMed)

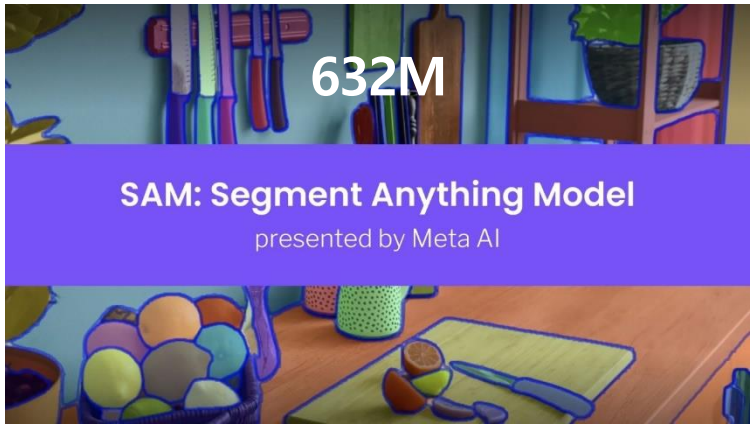
❖ SAMed 배경

- 자연 이미지 위주로 학습된 SAM은 의료 이미지에 적합하지 않음
- SAM을 Full Fine-tuning하는 것은 지나치게 큰 자원이 필요 → 효율적으로 Fine-tuning 해보자.



VS



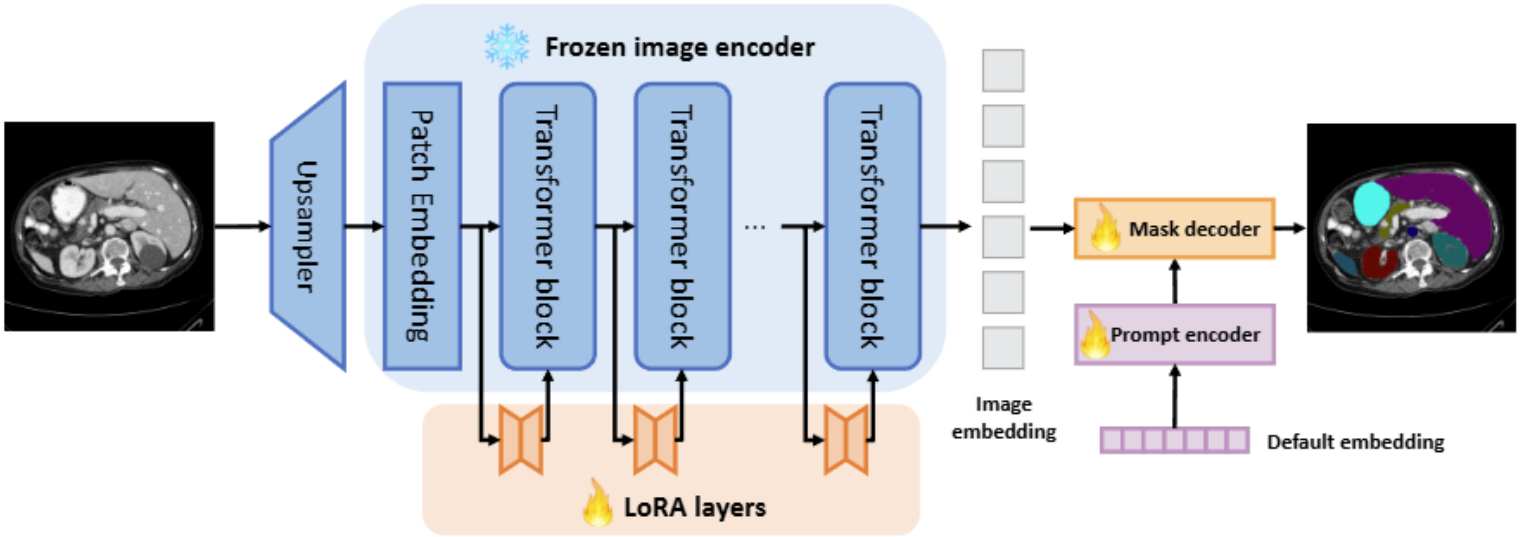


Algorithms

(3) Customized Segment Anything Model for Medical Image Segmentation (SAMed)

❖ SAMed 방법론

- SAM 내 전체 파라미터가 아닌, **일부 파라미터들만 튜닝**
- SAM Encoder에 의도적으로 추가한 Trainable Parameter & Mask Decoder & Prompt Encoder
  - ➢ Freeze: SAM Encoder 내 모든 파라미터 (95%)
  - ➢ Train: SAM Encoder에 의도적으로 추가한 Parameter & All Mask Decoder & All Prompt Encoder (5%)

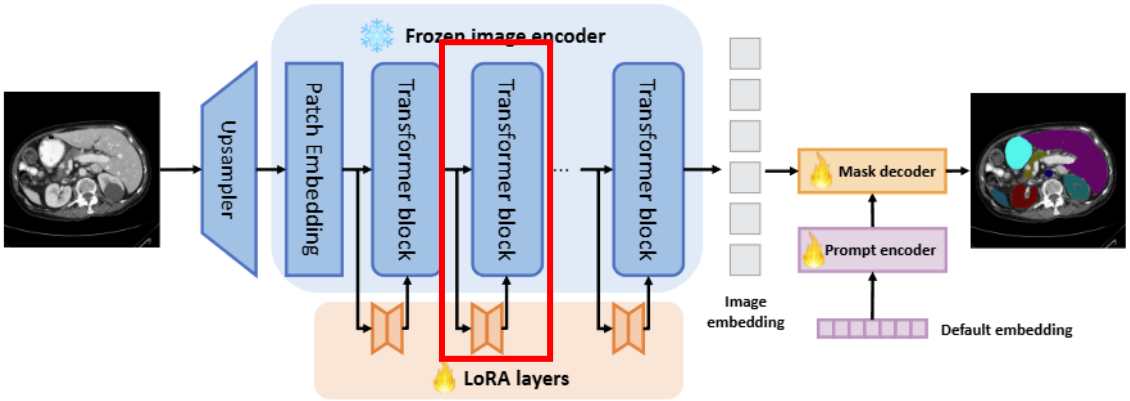


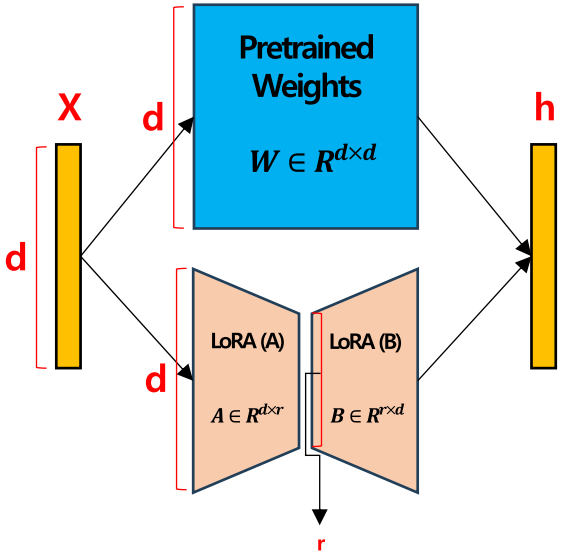
Algorithms

(3) Customized Segment Anything Model for Medical Image Segmentation (SAMed)

❖ SAMed 방법론 – SAM Encoder

- Encoder 내 기존 SAM 파라미터는 모두 Freeze **(기존 정보 유지)**
- 이때, Transformer Block 사이에 LoRA Layer를 추가하여, 해당 Layer만 학습 **(추가 정보 학습)**
  - ➢ Embedding 입력 → 기존 SAM Block 및 LoRA Layer에 통과 → 두 값을 합산 → LoRA Layer만 Update



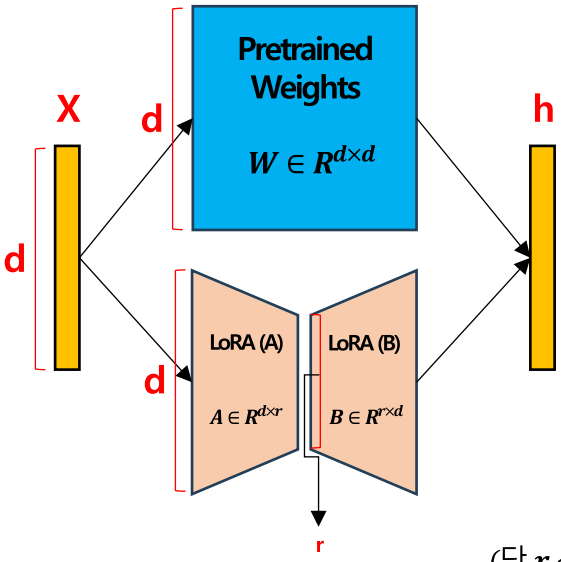


Algorithms

(3) Customized Segment Anything Model for Medical Image Segmentation (SAMed)

❖ SAMed 방법론 – SAM Encoder

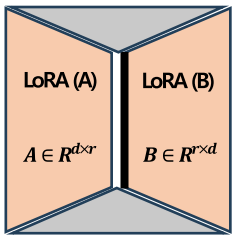
- LoRA (Low Rank Adaptation): AutoEncoder처럼 입력을 저차원으로 매핑 후 본 차원으로 확장
- SAM-Adapter보다 파라미터를 효율적으로 학습 가능
- Full Rank보다 노이즈 학습 가능성 최소화



(단, $r \ll d$)

Simple MLP (d^2개)

LoRA ($2dr$개)

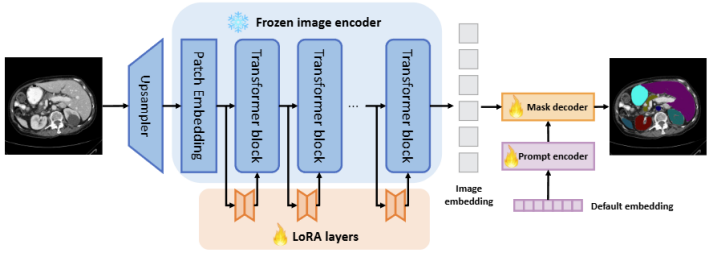


개 만큼 파라미터 개수가 효과적

저차원으로 매핑하기에, 불필요한 특징 학습 최소화

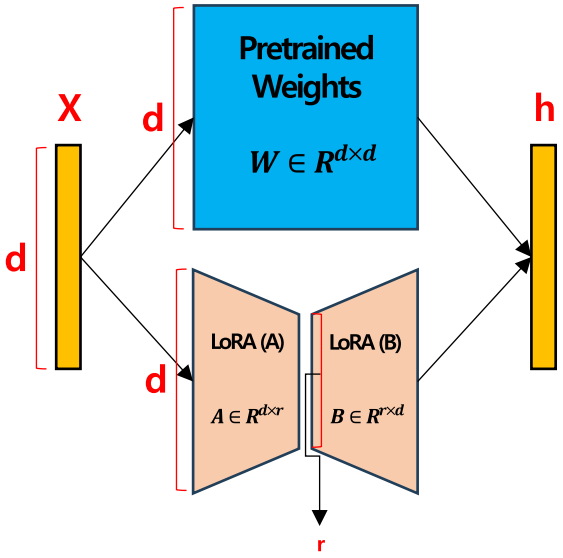
Algorithms

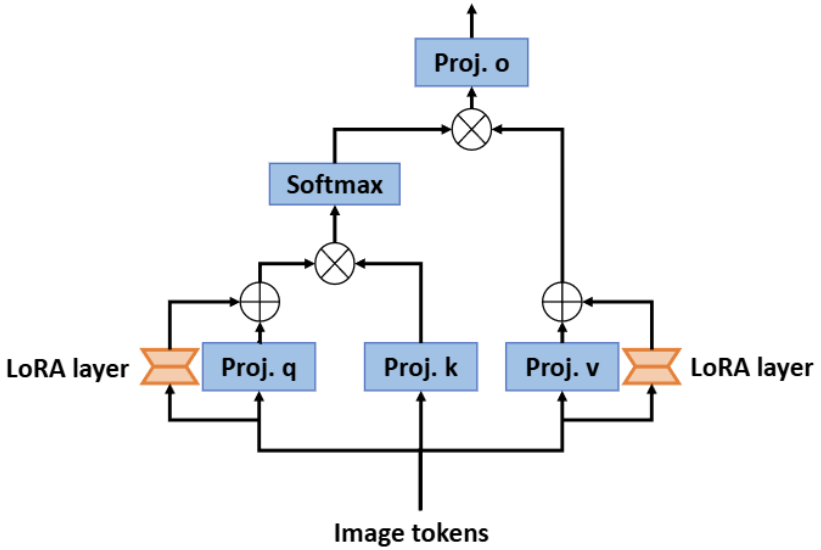
(3) Customized Segment Anything Model for Medical Image Segmentation (SAMed)



❖ SAMed 방법론 – SAM Encoder

- LoRA Layer는 Transformer Block 내 q, k, v, o 中 q와 v에서만 활용
  - ➢ k와 o는 별도 기존에 갖고 있던 Weight를 활용
  - ➢ 지나치게 Customize하여 SAM 본래 성능을 소실하는 것을 방지





$$Q = \hat{W}_q F = W_q F + B_q A_q F,$$
$$K = W_k F,$$
$$V = \hat{W}_v F = W_v F + B_v A_v F.$$

Algorithms

(3) Customized Segment Anything Model for Medical Image Segmentation (SAMed)

❖ SAMed 실험결과

- 기존 SOTA 모델들을 이기진 못했지만, 특정 데이터셋에서 SOTA에 버금가는 성능을 보여줌

Table 1: Quantitative comparison between SOTA methods and SAMed on the Synapse multi-organ CT dataset.

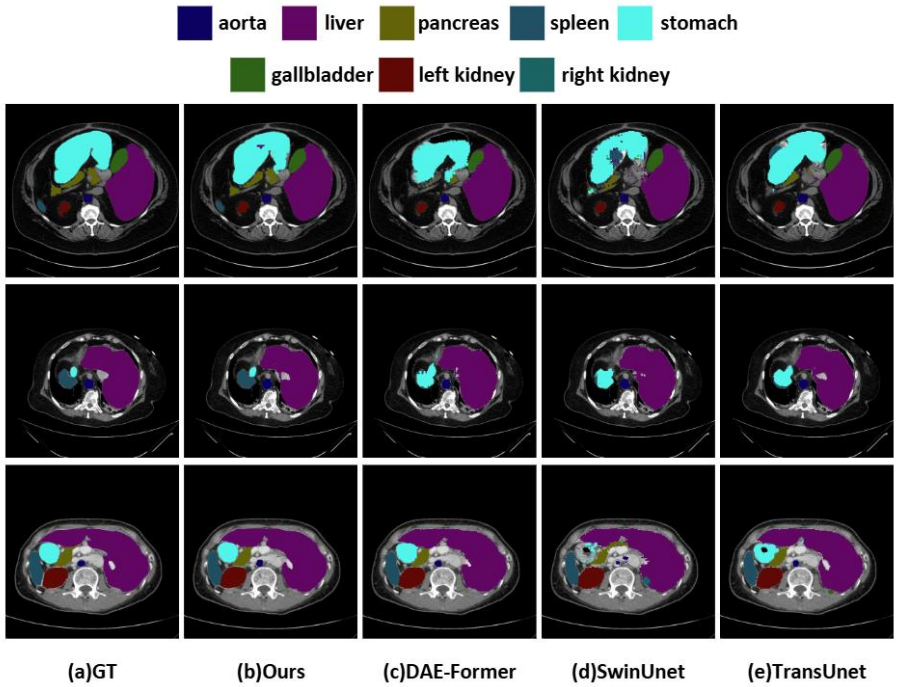
Methods	DSC↑	HD↓	Aorta	Gallbladder	Kidney(L)	Kidney(R)	Liver	Pancreas	Spleen	Stomach
U-Net [29]	76.85	39.70	89.07	69.72	77.77	68.60	93.43	53.98	86.67	75.58
Att-UNet [25]	77.77	36.02	**89.55**	68.88	77.98	71.11	93.57	58.04	87.30	75.75
TransUnet [4]	77.48	31.69	87.23	63.13	81.87	77.02	94.08	55.86	85.08	75.62
SwinUnet [3]	79.13	21.55	85.47	66.53	83.28	79.61	94.29	56.58	90.66	76.60
MissFormer [13]	81.96	18.20	86.99	68.65	85.21	**82.00**	94.41	65.67	91.92	80.81
TransDeepLab [2]	80.16	21.25	86.04	69.16	84.08	79.88	93.53	61.19	89.00	78.40
HiFormer [10]	80.39	**14.70**	86.21	65.69	85.23	79.77	94.61	59.52	90.99	81.08
DAE-Former [1]	**82.43**	17.46	88.96	**72.30**	**86.08**	80.88	**94.98**	65.12	**91.94**	79.19
SAMed	81.88	20.64	87.77	69.11	80.45	79.95	94.80	**72.17**	88.72	**82.06**

Algorithms

(3) Customized Segment Anything Model for Medical Image Segmentation (SAMed)

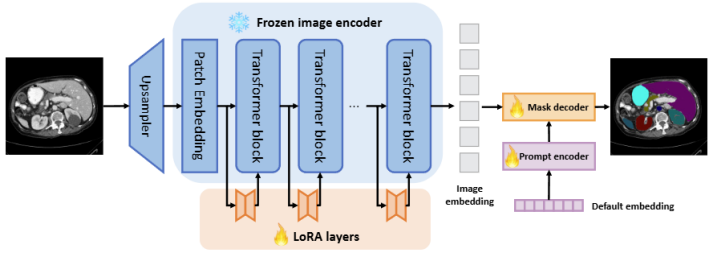
❖ SAMed 실험결과

- 기존 SOTA 모델들을 이기진 못했지만, 특정 데이터셋에서 SOTA에 버금가는 성능을 보여줌



Algorithms

(3) Customized Segment Anything Model for Medical Image Segmentation (SAMed)



❖ SAMed Ablation Study

- Encoder를 함께 학습 → 성능이 크게 개선
- Decoder에도 LoRA를 기반으로 학습 → Decoder는 LoRA 추가 부적합

Table 2: Ablation study on finetuning method of SAMed (The embedding in prompt encoder is finetuned by default).

Methods	DSC↑	Aorta	Gallbladder	Kidney(L)	Kidney(R)	Liver	Pancreas	Spleen	Stomach
Mask decoder	67.95	79.94	39.49	76.72	73.55	90.87	44.15	73.79	65.11
Image encoder + mask decoder	**81.88**	**87.77**	**69.11**	**80.45**	**79.95**	**94.80**	**72.17**	**88.72**	**82.06**

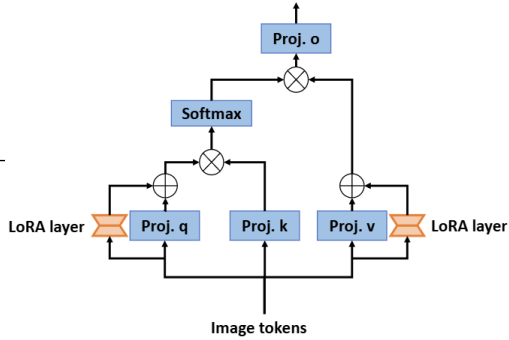
SAM은 자연 이미지에 초점 → Encoder를 함께 학습하지 않으면, 의료 이미지에 대한 효과적인 Feature 추출 불가

Table 3: Ablation study on applying LoRA finetuning to the transformer of mask decoder.

	Methods	DSC↑	Model size	Aorta	Gallbladder	Kidney(L)	Kidney(R)	Liver	Pancreas	Spleen	Stomach
Full Fine-tuning	SAMed	**81.88**	18.81M	**87.77**	**69.11**	**80.45**	**79.95**	**94.80**	**72.17**	**88.72**	**82.06**
LoRA	SAMed_s	77.78	**6.32M**	83.62	57.11	79.63	78.92	93.98	65.66	85.81	77.49

Algorithms

(3) Customized Segment Anything Model for Medical Image Segmentation (SAMed)



❖ SAMed Further Analysis: LoRA Layer

- LoRA에서 축소하는 Dimension이 4일 때, 가장 우수한 성능을 보임
- Q, K, V, O 中 Q와 V만 LoRA Layer를 붙였을 때 우수한 성능을 보임

Table 4: Ablation study on the rank size on the LoRA layer

Rank size	DSC↑	Aorta	Gallbladder	Kidney(L)	Kidney(R)	Liver	Pancreas	Spleen	Stomach
1	78.26	81.86	64.54	**81.97**	81.18	93.79	60.80	88.33	73.64
4	**81.88**	**87.77**	**69.11**	80.45	**79.95**	**94.80**	**72.17**	**88.72**	**82.06**
16	69.03	78.74	55.54	71.98	65.08	91.38	45.01	81.39	63.11

의료 이미지에 적합하기 위해서 학습은 필요 → But, 많은 학습 변수는 본래 SAM 기능을 저해 & 훈련 난이도 ↑

Table 5: Ablation study on the LoRA applied to different projection layers

Proj layer	DSC↑	Aorta	Gallbladder	Kidney(L)	Kidney(R)	Liver	Pancreas	Spleen	Stomach
Q	73.76	84.73	58.68	76.25	70.58	91.16	53.69	83.54	71.47
Q+V	**81.88**	**87.77**	**69.11**	**80.45**	**79.95**	**94.80**	**72.17**	**88.72**	**82.06**
Q+K+V+O	50.93	46.51	46.92	51.38	47.33	86.55	22.23	64.98	41.51

지나치게 많이 LoRA를 Customization → 본래 SAM 성능을 저하

Algorithms

(4) Medical sam adapter: Adapting segment anything model for medical image segmentation (Med-SA)

❖ **Medical sam adapter: Adapting segment anything model for medical image segmentation (Med-SA)**

- 2023년 arXiv에 수록 (인용수: 426회)
- **3D 의료 이미지에 적합하도록** Adapter를 활용하여 **SAM을 효율적으로 Fine-tuning**

Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation

Junde Wu[1,2,7], Wei Ji[3], Yuanpei Liu[8], Huazhu Fu[4], Min Xu[5,7], Yanwu Xu[6], Yueming Jin[2],

[1]University of Oxford, [2]National University of Singapore, [3]University of Alberta, [4]A*STAR, [5]Carnegie Mellon University, [6]Singapore Eye Research Institute, [7]Mohamed bin Zayed University of Artificial Intelligence, [8]University of Hongkong

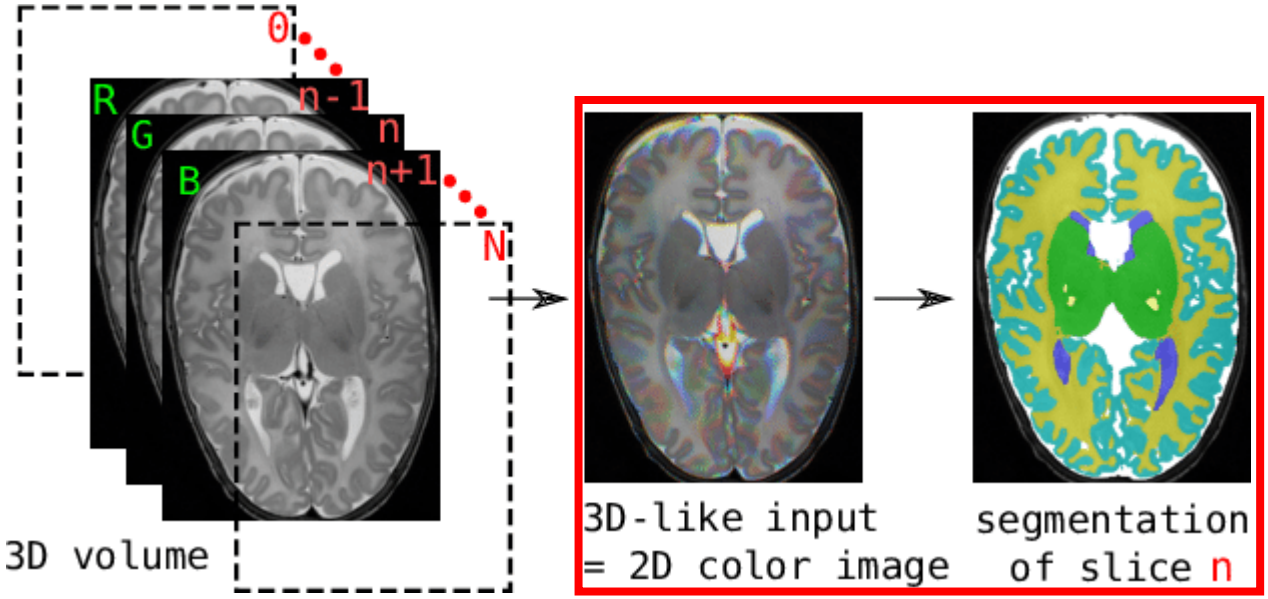
Wu, J., Ji, W., Liu, Y., Fu, H., Xu, M., Xu, Y., & Jin, Y. (2023). Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620.

Algorithms

(4) Medical sam adapter: Adapting segment anything model for medical image segmentation (Med-SA)

❖ Medical sam adapter 배경

- SAM이 의료 이미지에 대해서는 저조한 성능을 보임
- SAM을 효율적으로 Fine-tuning할 필요성
- **의료 이미지는 주로 3D인데, 3D 이미지에 맞도록 SAM Fine-tuning 할 수는 없을까?**

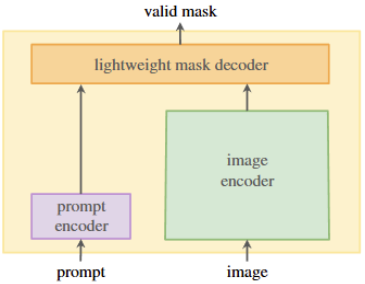


2D Segmentation은 3D 이미지 내 여러 세부 이미지들의 관계성을 충분히 반영하지 못함

[1] https://www.researchgate.net/publication/331085297/figure/fig1/AS:750836099596296@1556024684478/llustration-of-the-main-idea-used-in-15-a-segmentation-of-a-3D-medical-image-is.p

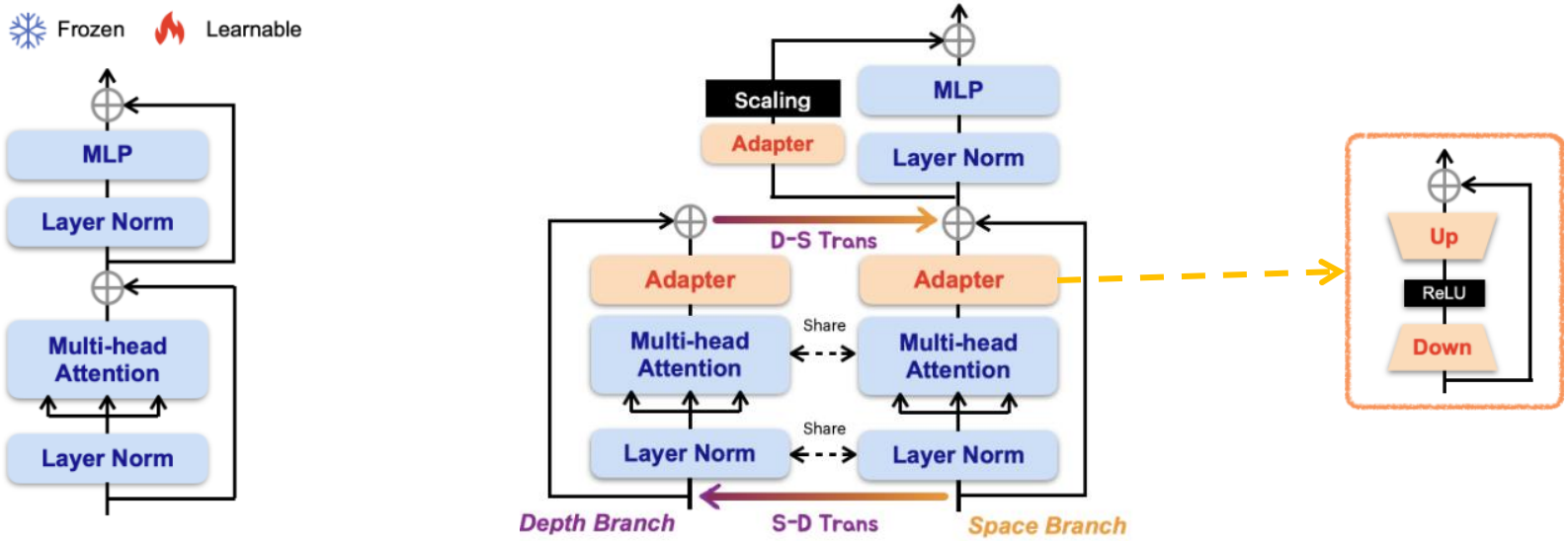
Algorithms

(4) Medical sam adapter: Adapting segment anything model for medical image segmentation (Med-SA)



❖ Medical sam adapter - Encoder

- Prompt Encoder: Freeze
- Image Encoder: 기본 SAM Weight는 Freeze & LoRA 기반 Adapter 3개 활용 & SD Trans 활용



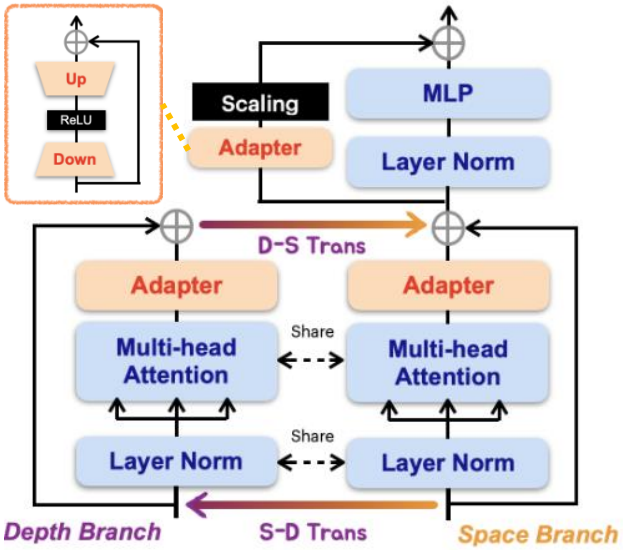
[Vanilla SAM] [Medical SAM Adapter] [Adapter]

Algorithms

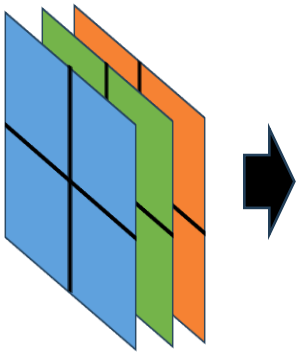
(4) Medical sam adapter: Adapting segment anything model for medical image segmentation (Med-SA)

❖ Medical sam adapter – Image Encoder

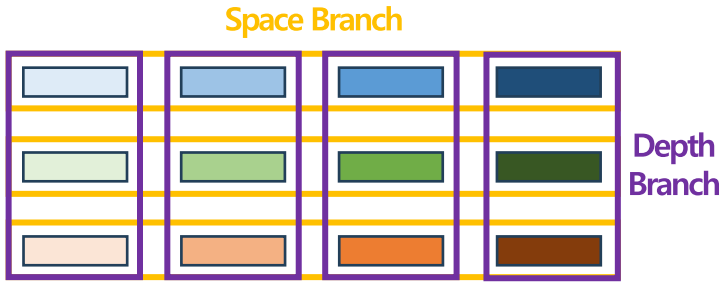
- LoRA 기반 Adapter: Full-rank MLP보다 효율적 & 노이즈에 강건한 학습 가능
- SD Trans: 2D에 적합한 SAM을 3D에 적합하기 위한 전략
  - ➢ 기존 Space 수준 Attention 뿐만 아니라, Depth 수준 Attention을 함께 학습



[이미지 단면1, 이미지 단면2, 이미지 단면3]



[이미지 단면1]



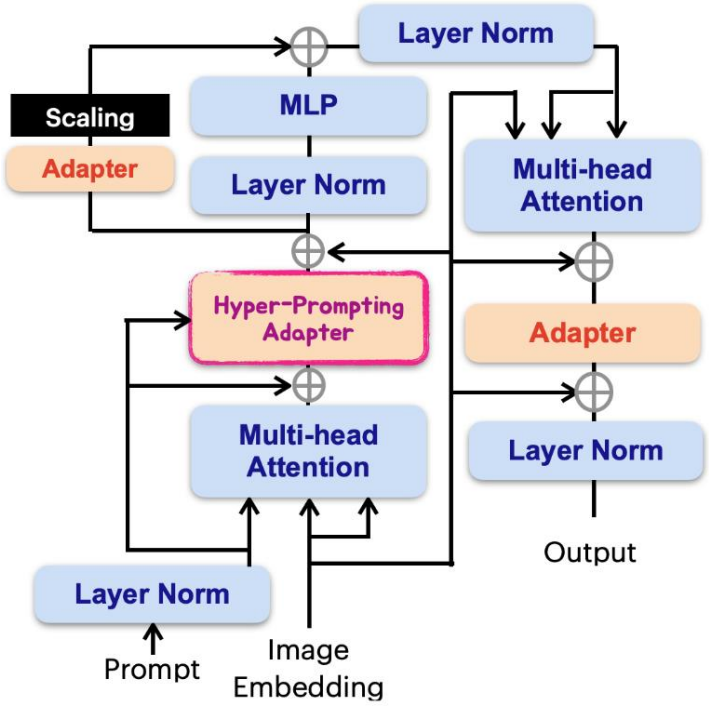
[각 이미지 내 패치 임베딩]

Algorithms

(4) Medical sam adapter: Adapting segment anything model for medical image segmentation (Med-SA)

❖ Medical sam adapter – Decoder

- Mask Decoder: $Embedding_{image}$와 $Embedding_{Prompt}$를 고려하여 적절한 Mask 산출
- 3개 Adapter를 Decoder에 삽입
  - ➢ Hyp-Adpt: $Embedding_{image}$와 $Embedding_{prompt}$의 결합을 강화하기 위한 Adapter

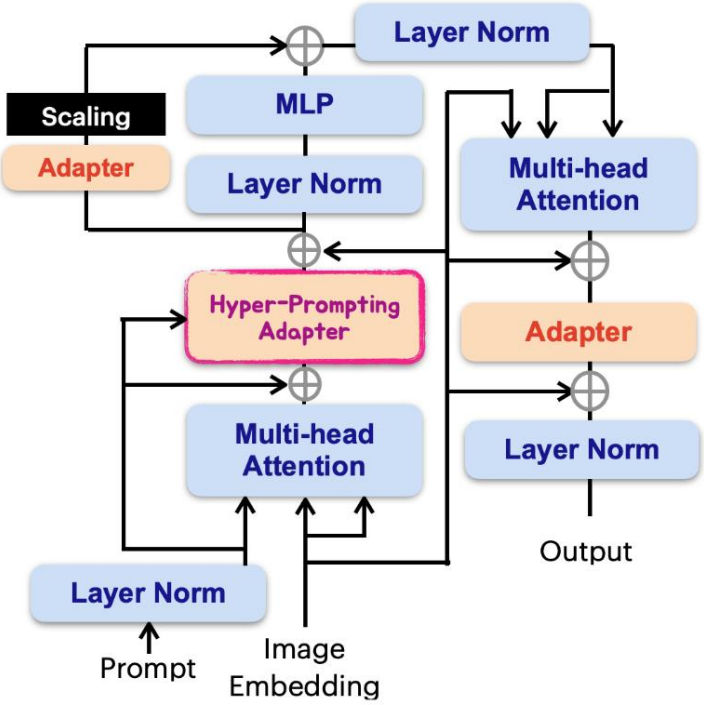


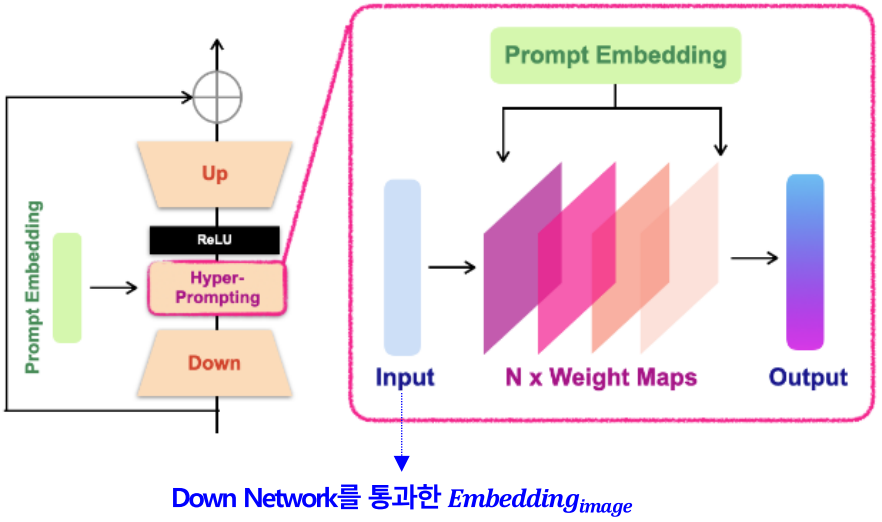
Algorithms

(4) Medical sam adapter: Adapting segment anything model for medical image segmentation (Med-SA)

❖ Medical sam adapter – Hyp-Adpt in Decoder

- HyperNetwork처럼, $Embedding_{prompt}$을 기반으로 Weight Matrix (W)생성
  - ➢ $W = Re\left(M\left(Embedding_{prompt}\right)\right)$: 이때, M은 Trainable MLP, Re는 Reshape를 의미)
- 생성된 Weight Matrix에 $Embedding_{image}$ 통과





Algorithms

(4) Medical sam adapter: Adapting segment anything model for medical image segmentation (Med-SA)

❖ Medical sam adapter 실험결과

- 기존 SAM 및 2D → 3D Adaption 방법론들보다 효과적인 성능을 보임
- 전체 미세조정보다, 일부(2%)만 Fine-tuning한 것이 효과적인 성능을 보임
- 프롬프트가 정교해질수록 우수한 성능을 보임 → 프롬프트 퀄리티 중요

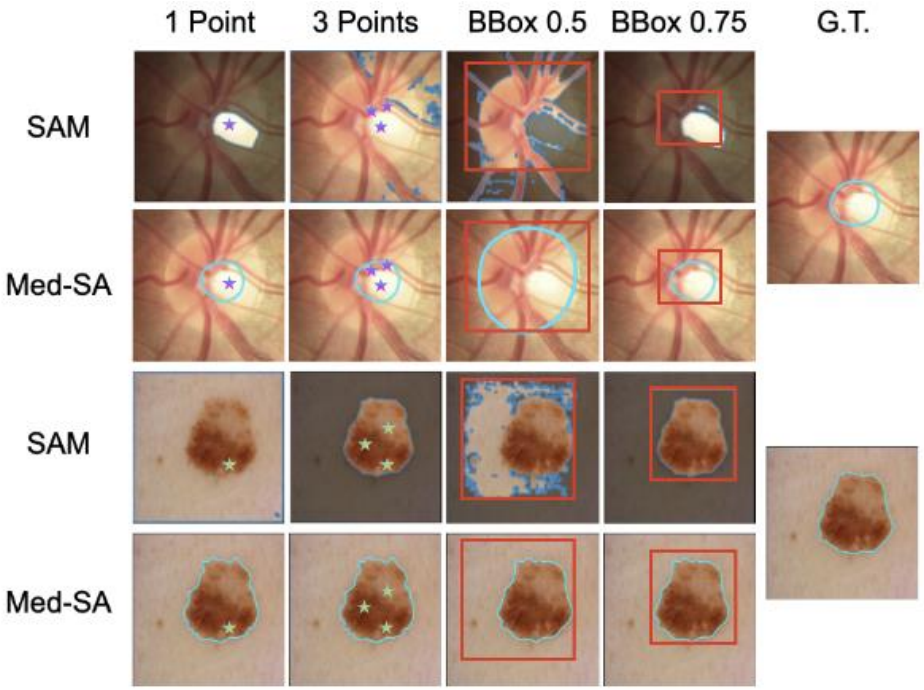
			Optic-Disc		Optic-Cup		Brain-Tumor			Thyroid Nodule		Melanoma	
	Param(M)	Turnable Param(M)	Dice	IoU	Dice	IoU	Dice	IoU	HD95	Dice	IoU	Dice	IoU
ResUNet	17	17	92.9	85.5	80.1	72.3	78.4	71.3	18.71	78.3	70.7	87.1	78.2
BEAL	25	25	93.7	86.1	83.5	74.1	78.8	71.7	18.53	78.6	71.6	86.6	78.0
TransBTS	39	39	94.1	87.2	85.4	75.7	87.6	78.44	12.44	83.8	75.5	88.1	80.6
EnsemDiff	32	32	94.3	87.8	84.2	74.4	88.7	80.9	10.85	83.9	75.3	88.2	80.7
MTSeg	27	27	90.3	83.6	82.3	73.1	82.2	74.5	15.74	82.3	75.2	87.5	79.7
UltraUNet	19	19	91.5	82.8	83.1	73.8	84.5	76.3	14.03	84.5	76.2	89.0	81.8
FAT-Net	75	75	91.8	84.8	80.9	71.5	79.2	72.8	17.35	80.8	73.4	90.7	83.9
BAT	88	88	92.3	85.8	82.0	73.2	79.6	73.5	15.49	81.7	74.2	91.2	84.3
SegDiff	32	32	92.6	85.2	82.5	71.9	85.7	77.0	14.31	81.9	74.8	87.3	79.4
nnUNet	16	16	94.7	87.3	84.9	75.1	88.5	80.6	11.20	84.2	76.2	90.8	83.6
TransUNet	96	96	95.0	87.7	85.6	75.9	86.6	79.0	13.74	83.5	75.1	89.4	82.2
UNetr	104	104	94.9	87.5	83.2	73.3	87.3	80.6	12.81	81.7	73.5	89.7	82.8
Swin-UNetr	138	138	95.3	87.9	84.3	74.5	88.4	81.8	11.36	83.5	74.8	90.2	83.1
SAM 1 points	636	0	-	-	-	-	63.2	47.6	32.53	-	-	81.6	70.4
SAM 3 points	636	0	-	-	-	-	71.3	64.5	28.74	-	-	85.8	77.5
SAM BBox 0.5	636	0	-	-	-	-	51.2	44.6	38.56	-	-	75.3	64.8
SAM BBox 0.75	636	0	-	-	-	-	74.6	62.1	27.51	-	-	85.7	74.4
MedSAM 1 point	636	636	92.9	85.5	82.1	73.8	81.5	74.3	15.68	81.3	74.7	86.8	77.5
MedSAM 3 points	636	636	93.8	86.2	82.8	74.2	82.3	74.8	15.19	81.6	75.1	87.5	78.6
MedSAM BBox 0.5	636	636	92.6	85.3	82.0	75.2	82.0	74.7	15.05	82.4	75.5	88.5	79.2
MedSAM BBox 0.75	636	636	94.6	86.7	82.8	75.9	83.6	75.6	14.90	82.8	75.7	88.9	79.8
Med-SA 1 point	636	13	97.4	89.5	86.8	78.8	89.1	81.8	10.38	86.3	78.7	92.6	84.1
Med-SA 3 points	636	13	97.9	89.8	87.1	79.0	89.8	82.3	10.11	86.7	79.4	**93.4**	**84.7**
Med-SA BBox 0.5	636	13	97.6	89.6	86.4	78.5	89.5	81.9	10.35	86.6	78.9	92.1	83.0
Med-SA BBox 0.75	636	13	**98.3**	**90.1**	**87.5**	**79.9**	**90.5**	**83.0**	**9.50**	**88.4**	**80.4**	93.0	84.2

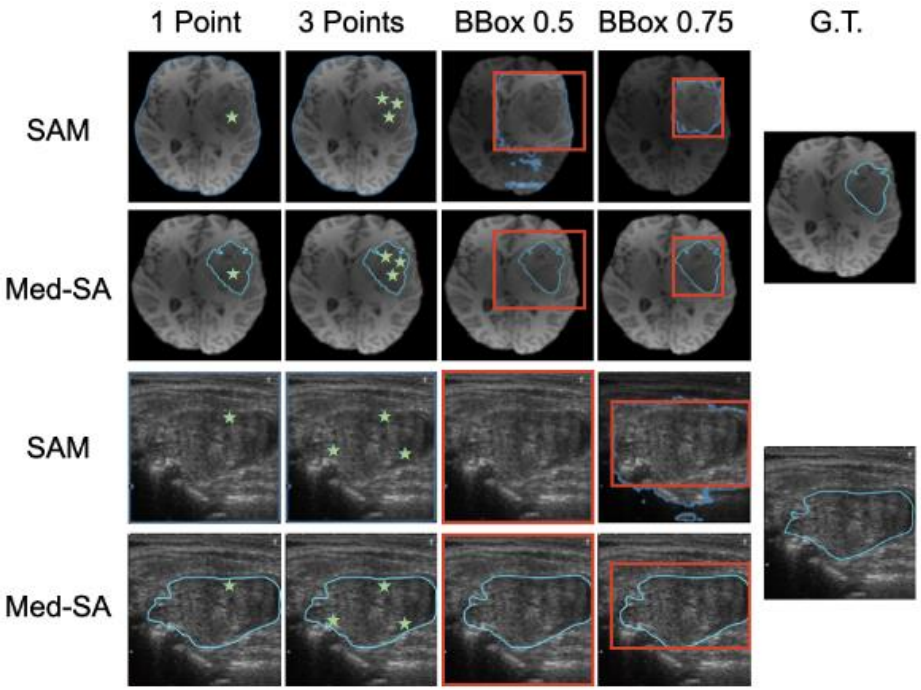
Algorithms

(4) Medical sam adapter: Adapting segment anything model for medical image segmentation (Med-SA)

❖ Medical sam adapter 실험결과

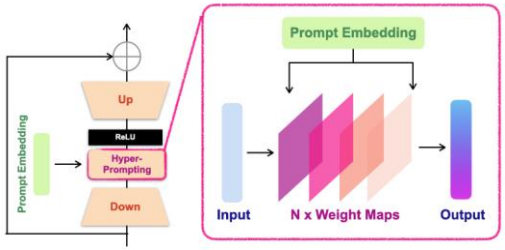
- SAM은 프롬프트 편차가 매우 큼 → 정교한 프롬프트 없이는 사실상 이용불가
- Medical sam adapter는 프롬프트가 다소 불안정하더라도, GT와 유사한 성능을 보임





Algorithms

(4) Medical sam adapter: Adapting segment anything model for medical image segmentation (Med-SA)



❖ Medical sam adapter Ablation Study

- SD-Trans가 추가됨으로써 우수한 성능을 보임
- Prompt Condition을 주는 방법들 중 Hyp-Adpt가 가장 좋은 성능을 보임
  - ➢ Add: $Embedding_{prompt} + Embedding_{down}$
  - ➢ Concat: $[Embedding_{prompt}, Embedding_{down}]$
  - ➢ Hyp-Adpt: $Embedding_{prompt}$로 만든 Network에 $Embedding_{down}$을 통과시킴

2D-3D	Prompt-Condition			BTCV	OpticDisc	OpticCup	BrainTumor	ThyroidNodule	Melanoma
SD-Trans	Add	Concat	**HyP-Adpt**	Ave-Dice (%)	Dice (%)	Dice (%)	Dice (%)	Dice (%)	Dice (%)
				79.3	90.1	80.1	77.5	76.5	89.2
✓				84.7	-	-	81.7	-	-
✓	✓			86.1	94.6	83.4	83.9	83.7	93.8
✓		✓		86.4	95.7	84.0	85.1	84.8	94.5
✓			✓	**88.3**	**97.4**	**86.8**	**87.6**	**86.3**	**96.3**

Conclusion

Conclusion

① 632M 파라미터의 거대 모델 구조

② 약 10억개 Mask로 학습

⬇

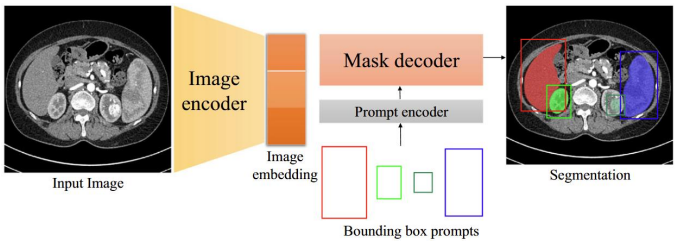
추가적인 학습 없이 일반적인 객체에 우수한 Segmentation 모델

의료 이미지 등 특수한 데이터에는 불완전한 Segmentation 성능을 보임

Q. 어떻게 SAM을 이러한 특수한 데이터에 적용할 수 있을까?

① SAM을 Full Fine-tuning

모든 Encoder & Decoder Weight를 학습



MedSAM: 많은 의료 이미지로 SAM을 모두 학습

① 학습 시, 많은 자원이 필요함

② 잘 학습된 SAM의 기존 지식 활용 불가

② SAM을 Parameter Efficient Fine-tuning

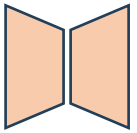
Encoder 내 부가적인 Weight & 모든 Decoder Weight를 학습

(1) SAM-Adapter

- Encoder 중간에 삽입한 MLP Adapter

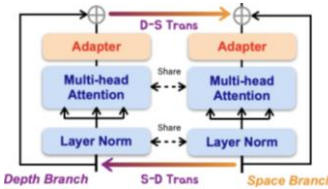
(2) SAMed

- Down & Up 형태의 LoRA Adapter
- MLP보다 효율적이고, 노이즈에 강건



(3) Medical sam adapter

- LoRA 기반 Adapter로 SAM을 Fine-tuning
- SD-Trans로 3D 이미지에 적합하도록 학습



Reference

1. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4015-4026).
2. Ji, W., Li, J., Bi, Q., Liu, T., Li, W., & Cheng, L. (2024). Segment anything is not always perfect: An investigation of sam on different real-world applications. Machine Intelligence Research.
3. Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment anything in medical images. Nature Communications, 15(1), 654.
4. Chen, T., Zhu, L., Deng, C., Cao, R., Wang, Y., Zhang, S., ... & Mao, P. (2023). Sam-adapter: Adapting segment anything in underperformed scenes. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3367-3375).
5. Zhang, K., & Liu, D. (2023). Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785.
6. Wu, J., Ji, W., Liu, Y., Fu, H., Xu, M., Xu, Y., & Jin, Y. (2023). Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620.

Thank you!