Medical Image Segmentation Based on Large Models

第13组：韩轾 刘翰文 吴竞尧 姜楠

上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

目 录
CONTENT

上海交通大學

SHANGHAI JIAO TONG UNIVERSITY

PART ONE

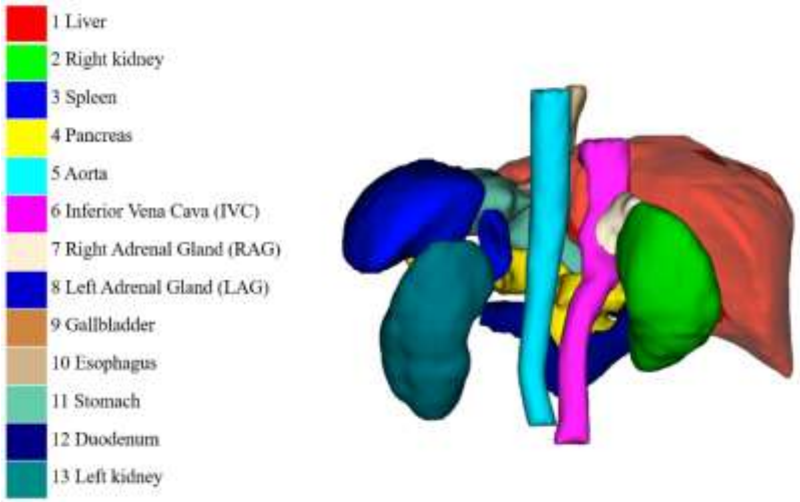
Introduction

Introduction

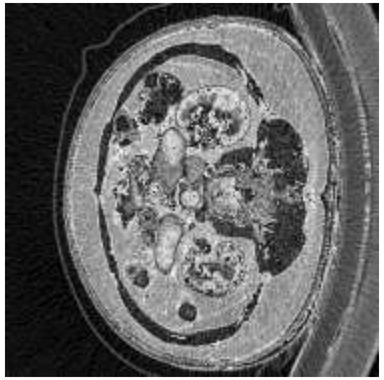
FLARE22 Dataset Info

FLARE22 is a specialized dataset for medical image segmentation, designed to promote the development of federated learning and adaptive tuning techniques, which includes segmentation masks for 13 anatomical structures, providing a rich set of training and testing samples to facilitate the evaluation of models in various medical imaging tasks.

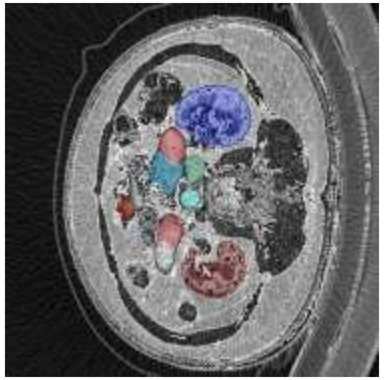
FLARE22 uses 2D slices from 3D volumetric objects. The 2D form of image-mask pairs significantly reduces computational complexity and memory usage, making training and inference more efficient and compatible with most existing models.



Segmentation(0-13)



image



mask

Introduction

Experimental Methods

Primary experiments focus on evaluating the performance of some models on FLARE22 dataset:

Zero-shot

- SAM & SAM2
- MedSAM

Fine-tune

- LoRA on SAM
- SAM Adapter

In addition, we investigated the performance of some models on downstream scenario datasets:

- SAM & SAM2
- LoRA on SAM

Downstream Scenario: Remote-sensing & Infrared

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

PART TWO

Zero-Shot Evaluation

SJTU

SAM/SAM2

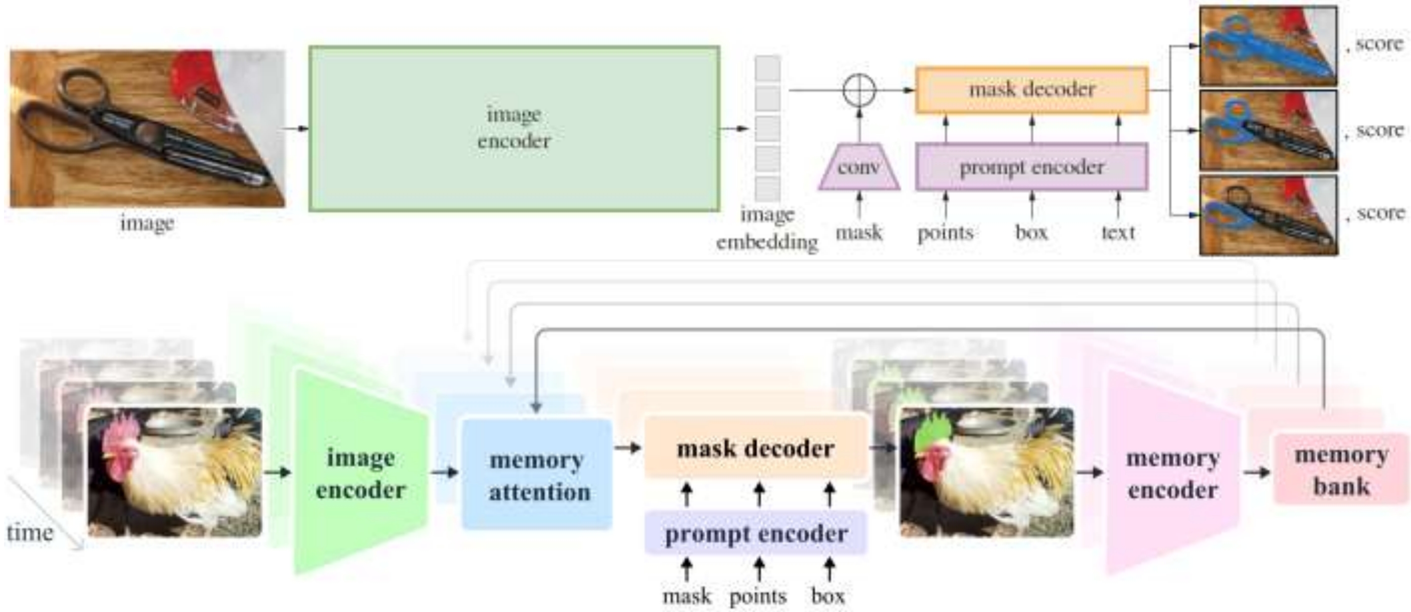
SAM / SAM2 Info

SAM

- General Segmentation based on ViT
- Prompt-based Segmentation
- Multi-scale Feature Fusion

SAM 2

- Memory Structure for spatiotemporal processing
- Accepting points, box, or mask prompts.



SAM

SAM 2

SAM/SAM2

Comparison of Zero-Shot Results

SAM outperformed SAM 2 in the zero-shot segmentation results for all 13 anatomical structures when applying to FLARE22 Dataset.

Reasoning:

- 2D medical imaging data lack spatiotemporal information and contain less information overall.
- SAM 2 contains more complex hyperparameters, which may lead to a decrease in generalization ability.

Table 3. Accuracy of SAM and SAM2 on Flare

Organs	SAM	SAM2
background	0.4678	0.3357
liver	0.5418	0.4255
right kidney	0.7789	0.629
spleen	0.8002	0.636
pancreas	0.2321	0.0647
aorta	0.7915	0.5825
inferior vena Cava(IVC)	0.4659	0.2915
right adrenal gland	0.0253	0.0081
left adrenal gland	0.1891	0.038
gallbladder	0.4196	0.152
esophagus	0.2759	0.0609
stomach	0.4558	0.2035
duodenum	0.2937	0.1385
left kidney	0.8049	0.6962
Average	0.4673	0.3044

上海交通大學

SHANGHAI JIAO TONG UNIVERSITY

PART THREE

Finetuning on SAM

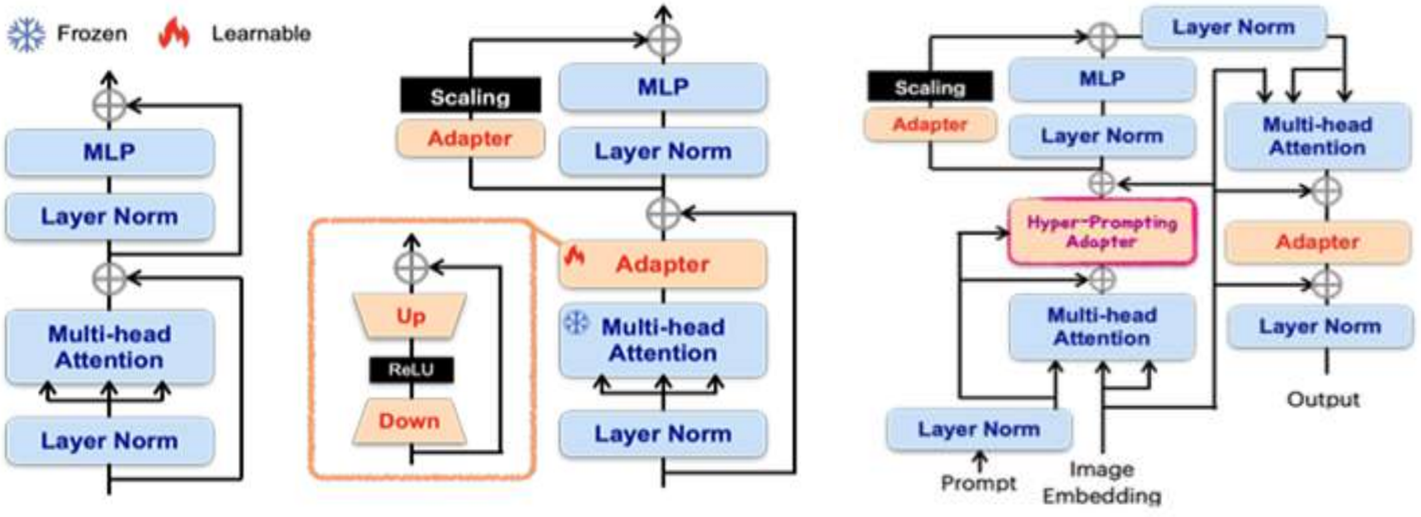
Finetuning

Adapter ‘Junde Wu, Wei Ji, Yuanpei Liu, Huazhu Fu, Min Xu, Yanwu Xu, and Yueming Jin. Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620, 2023.’

Image encoder: adapter after multi-head attention, residual path of MLP

Prompt encoder: full fine tuning with default embedding

Mask Decoder: hyper-prompting adapter(not used), adapter after multi-head attention, residual path of MLP



a. ViT Block in SAM

b. 2D Medical Image Adaption

d. Decoder Adaption with prompt

Finetuning

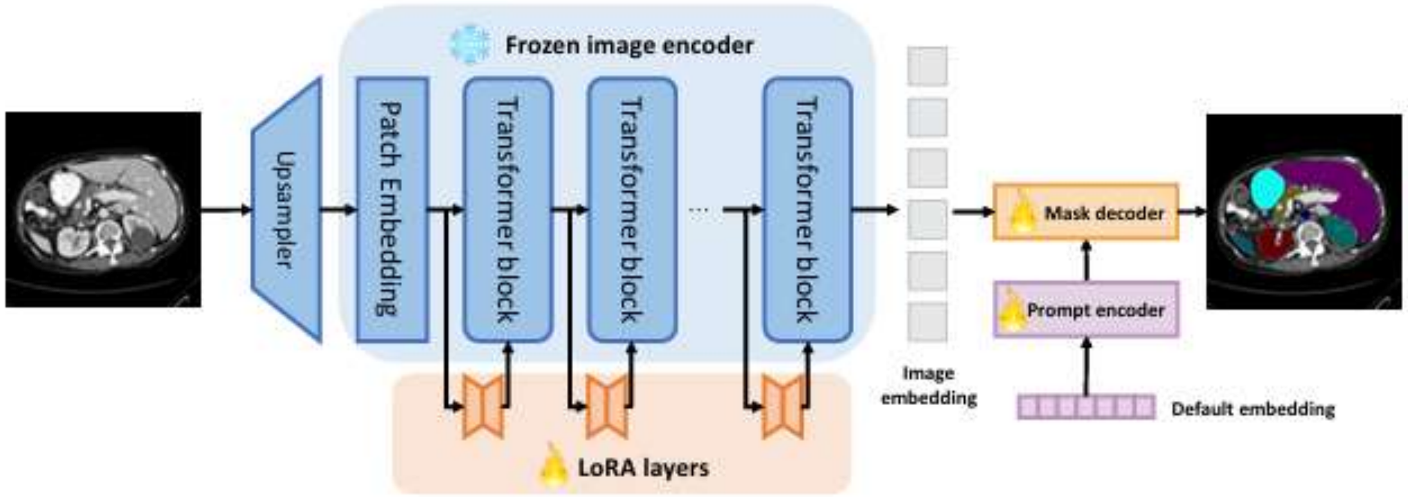
LoRA 'Kaidong Zhang and Dong Liu. Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785, 2023.'

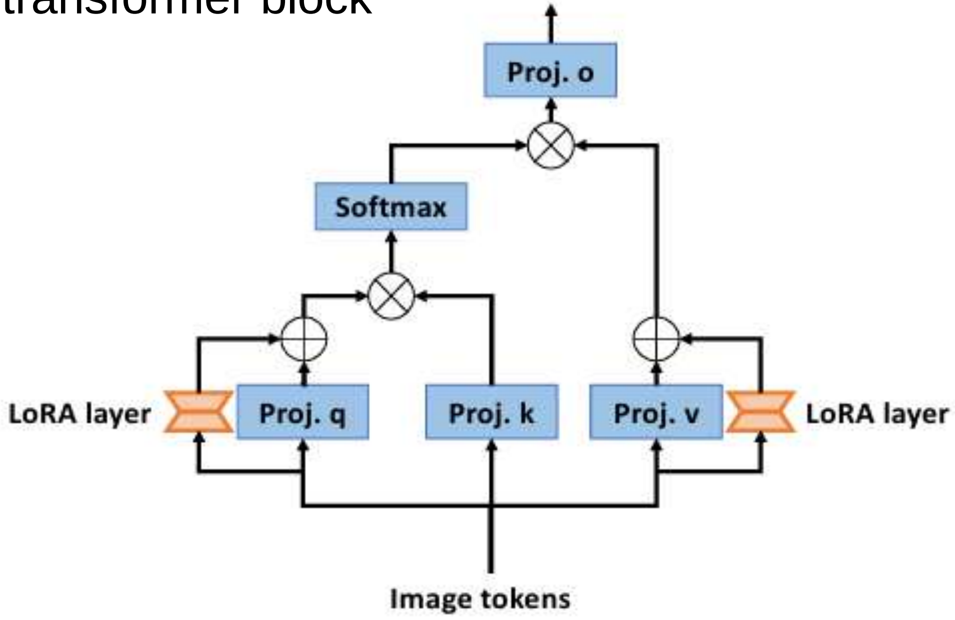
Image encoder: apply LoRA layer to the **Q** and **V** projection layers of each of the transformer block

$$\hat{W} = W + \Delta W = W + BA$$

Prompt encoder: full fine tuning with default embedding

Mask Decoder: LoRA to **Q** and **V** projection layers of each of the transformer block





Finetuning

Ablation Study

- For the LoRA fine-tuned SAM, the version with rank r=4 performed best.
- LoRA generally outperformed Adapter to some extent.

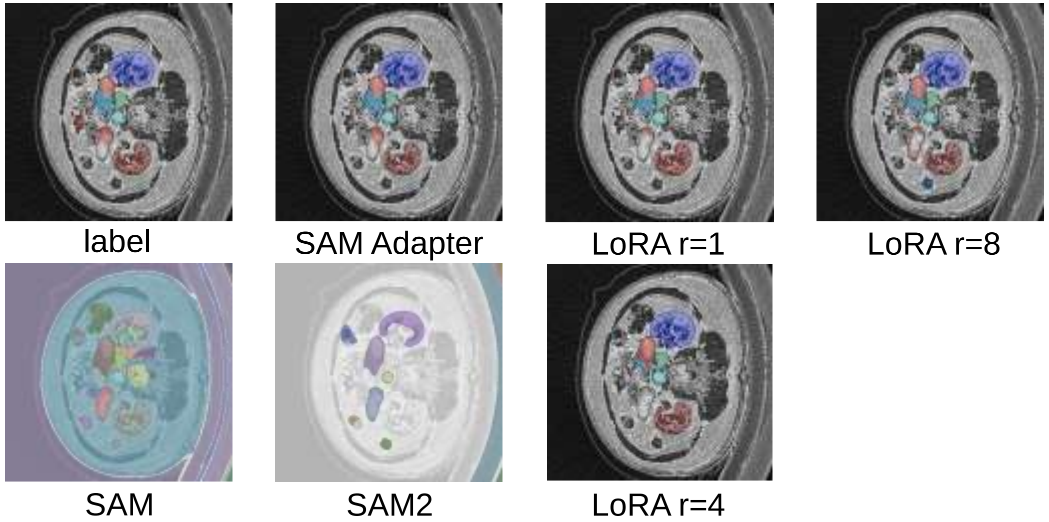


Table 2. Accuracy of SAM, SAM Adapter, Sam Lora(rank=1, 4, 8) on Flare

Organs	SAM	SAM Adapter	SAM Lora(rank=1)	SAM Lora(rank=4)	SAM Lora(rank=8)
background	0.4678	0.9973	0.9944	0.9961	0.9953
liver	0.5418	0.7328	0.6758	0.7191	0.7093
right kidney	0.7789	0.5536	0.526	0.5224	0.4828
spleen	0.8002	0.413	0.3901	0.4406	0.4417
pancreas	0.2321	0.4307	0.3826	0.401	0.3985
aorta	0.7915	0.7985	0.7759	0.7975	0.7872
inferior vena Cava(IVC)	0.4659	0.6611	0.5818	0.6342	0.6309
right adrenal gland	0.0253	0.1777	0.1778	0.1914	0.185
left adrenal gland	0.1891	0.1993	0.1761	0.1874	0.1538
gallbladder	0.4196	0.2674	0.2546	0.2802	0.2578
esophagus	0.2759	0.353	0.3368	0.365	0.3654
stomach	0.4558	0.6076	0.627	0.6897	0.6554
duodenum	0.2937	0.7392	0.6792	0.7519	0.7192
left kidney	0.8049	0.9367	0.9265	0.958	0.9333
Average	0.4673	0.5620	0.5360	0.5667	0.5511

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

PART FOUR

Downstream Scenario Implementation

SJTU

Finetuning

Remote Sensing

- Dataset:
  - [GID]
- Mask Categories:
  - Buildings
  - Water
  - Woods
  - Lawn
  - Farmland

- Segmentation Results:
  - LoRA Fine-Tuned SAM showed improved performance on most mask categories compared to the zero-shot version.

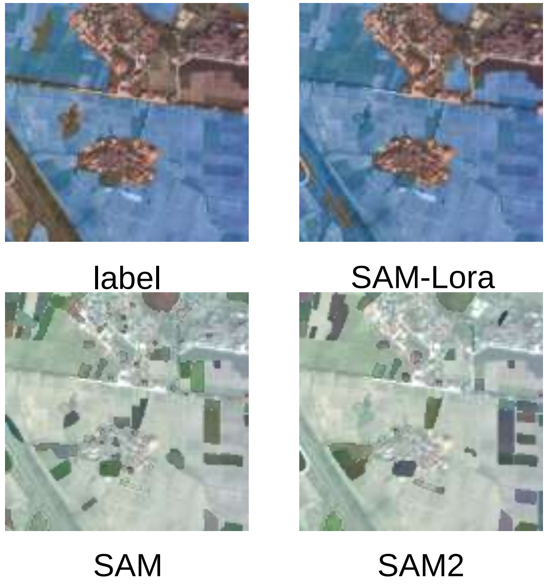


Table 4. Accuracy of SAM, SAM2 and SAM LORA on remote sensing datasets

Classification	SAM	SAM2	SAM Lora
background	0.1059	0.0975	0.8047
buildings	0.1742	0.1385	0.6646
water	0.2771	0.3197	0.4524
woods	0.8107	0.8027	0.3568
lawn	0.8367	0.8249	0.3673
farmland	0.2435	0.2325	0.6411
Average	0.4080	0.4026	0.5478

Finetuning

Infrared

- Dataset:
  - [SIRST-5K]
- Mask Categories:
  - Target / Background
- Segmentation Results:
  - The LoRA fine-tuned SAM showed a significant improvement in segmentation performance.


label


SAM-Lora


SAM


SAM2

Table 5. Accuracy of SAM, SAM2 and SAM LORA on infrared datasets

Classification	SAM	SAM2	SAM Lora
background	0.7614	0.4407	0.9999
detect target	0.5203	0.4379	0.8874
Average	0.64085	0.4393	0.94365

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

THANKS FOR WATCHING

SJTU