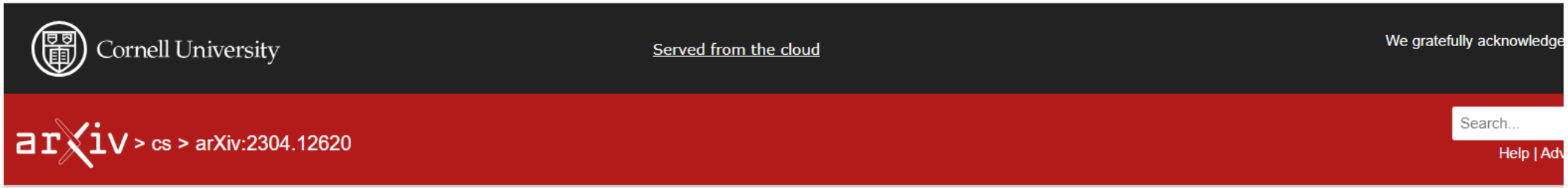
Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation

Yu Jiening
Yu Jiening@umac.mo

um 澳大



arXiv preprint arXiv:
2304.12620, 2023.

Catalogue

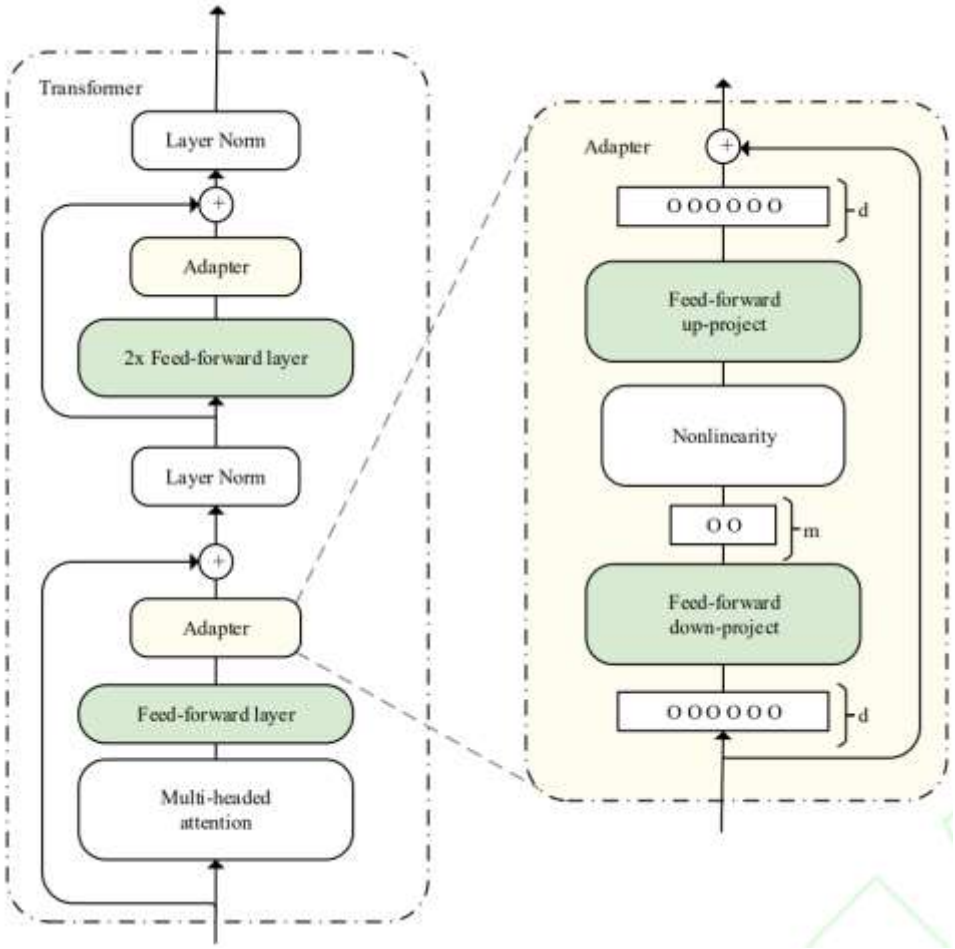
- Introduction
- Method
- Experiments
- Conclusion

1.Introduction

- Fine-tuning SAM to enable its use in medical images.
- Choose to fine-tune the pre-trained SAM using a **parameter-efficient fine-tuning (PEFT)** technique called **Adaption**.
- The main idea is to **insert several parameter-efficient Adapter** modules into the original fundamental model, and then **only adjust the Adapter parameter** while leaving all pre-trained parameters frozen.

Why PEFT and Adaption?

- PEFT learns fewer parameters than full fine-tuning, making efficient learning possible.
- Among all PEFT strategies, Adaption can be easily adopted in various downstream computer vision tasks.

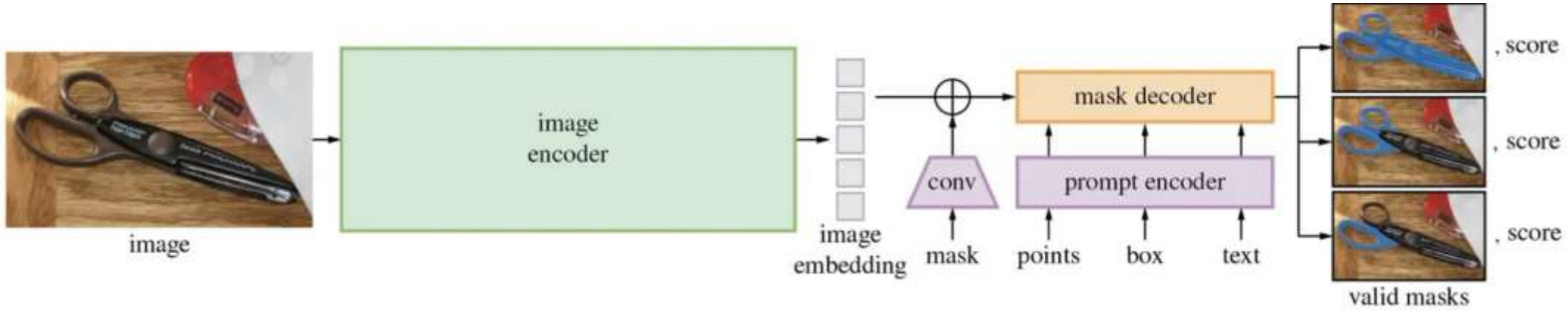


Contributions

- Extending the SAM model to the medical domain.
- The first to propose the adaptation approach for the general medical image segmentation.
- We have evaluated our proposed **MSA model** on **19 medical image segmentation tasks** with different image modalities including MRI, CT, fundus image, ultrasound image, and dermoscopic images. Our results demonstrate that MSA outperforms the previous state-of-the-art methods by a considerable margin.

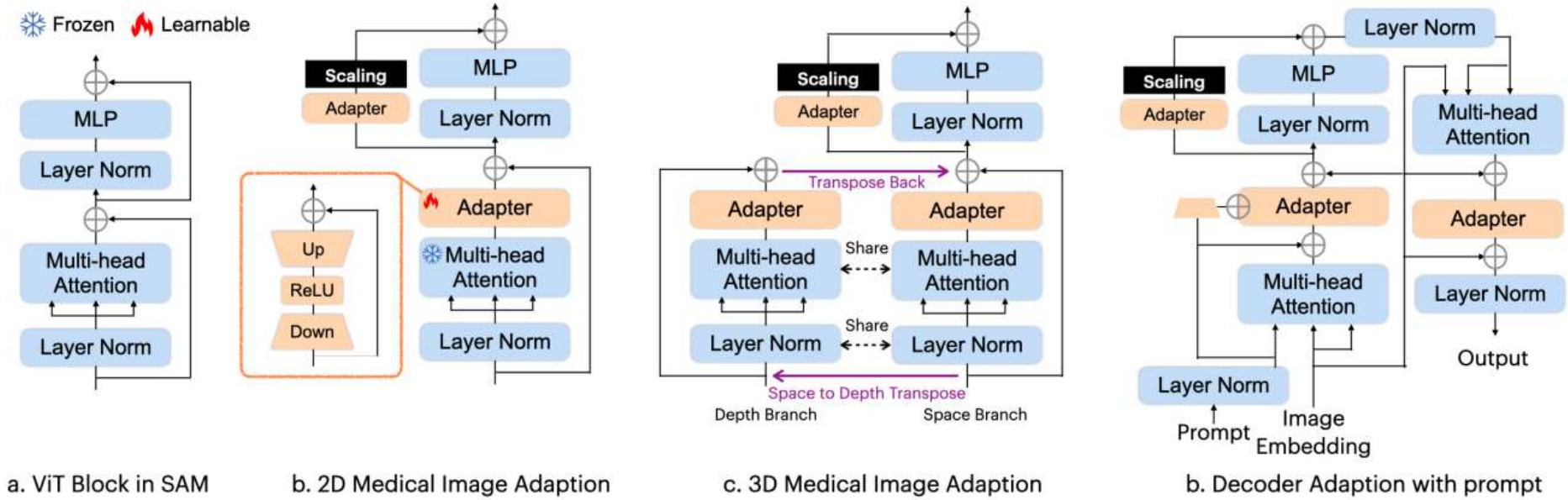
2.Method

2.1 Preliminary: SAM architecture



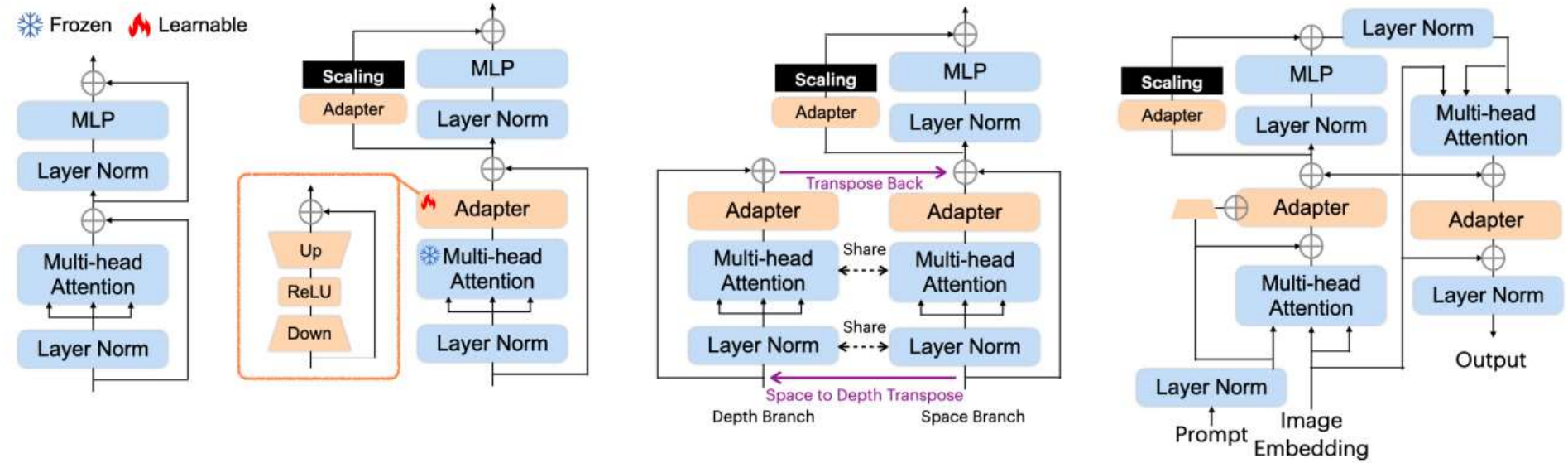
- SAM comprises three main components: an image encoder, a prompt encoder, and a mask decoder.

2.2 MSA architecture



a. ViT Block in SAM b. 2D Medical Image Adaption c. 3D Medical Image Adaption b. Decoder Adaption with prompt

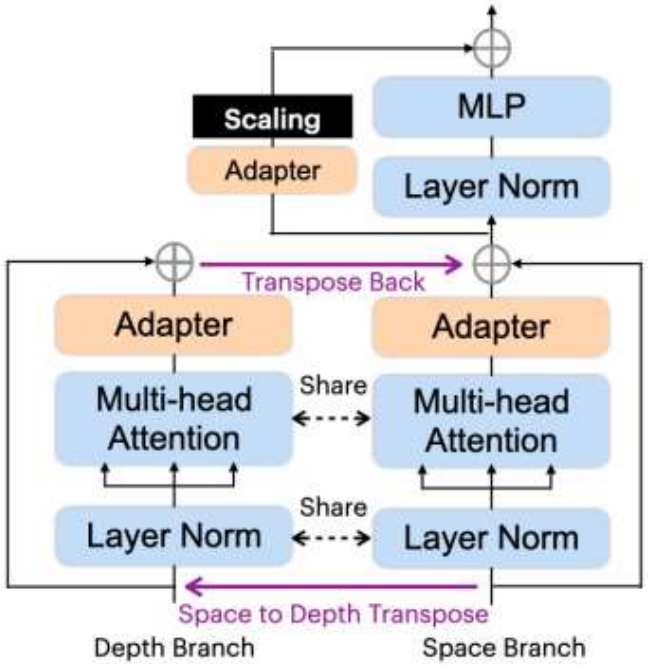
- In SAM encoder ,place **the first Adapter** after the multi-attention and before the residual connection, and **the second Adapter** in the residual shortcut of the MLP after the multi-attention.



a. ViT Block in SAM b. 2D Medical Image Adaption c. 3D Medical Image Adaption b. Decoder Adaption with prompt

- In the SAM decoder, three Adapters are deployed per ViT block.
- the first one is placed after the multiple cross attention from the prompt to the image embedding with residual summation of the prompt embedding
- the second one is deployed in exactly the same way as the encoder
- the third one is deployed after the residual concatenation of the image embedding vectors

- To address the correlation in the depth dimension of 3D medical images, the attention operation is divided into **space branch** and **depth branch**.
- Send **D × N × L** to the polytopic attention of the space branch and transpose the input matrix on the depth branch to obtain an **N × D × L** input.
- N:the number of embedding vectors
- L: the length of the embedding vectors
- D:the number of operations



c. 3D Medical Image Adaption

2.3 Training Strategy

Encoder Pre-training:

- Use a mixture of four medical image datasets for this pretraining, including the **RadImageNet** dataset, **EyePACSp** dataset, the **BCN-20000** and **HAM-10000** datasets.
- **RadImageNet** is a large-scale collection of 1.35 million radiological images (CT, MRI, US) covering a wide range of organs such as ankles/feet, brain, hips, knees, shoulders, spine, abdomen, pelvis, chest, pelvis and thyroid.
- The **EyePACSp** dataset contains 88,702 colour fundus images.
- The **BCN-20000** and **HAM-10000** contain approximately 30,000 dermoscopic images with melanoma or nevi on the image.

Prompt training:

- Adapted SAM on the new dataset. The process is essentially the same as in SAM, but with the following modification: for clicked cues, **positive clicks indicate foreground regions and negative clicks indicate background regions**.
- It is first initialised using random sampling and then some click operations are added using an iterative sampling process.

3.Experiments

3.1 Dataset

- Conducted experiments on **five different medical image** segmentation datasets, categorized into **two types**.
- The first type tested the **overall segmentation performance** of the model, comparing it to commonly used medical segmentation **baselines** and **state-of-the-art (SOTA)** methods.
- The other four tasks were used to verify the model's generalization to different modalities and kinds of tasks

3.2 Main Results

The comparison of MSA with SOTA segmentation methods and original SAM over AMOS dataset evaluated by Dice Score.

Methods	Spleen	R.Kid	L.Kid	Gall.	Eso.	Liver	Stom.	Aorta	IVC	Panc.	RAG	LAG	Duo.	Blad.	Postc.	Avg
TransUNet	0.881	0.928	0.919	0.813	0.740	0.973	0.832	0.919	0.841	0.713	0.638	0.565	0.685	0.748	0.692	0.792
EnsDiff	0.905	0.918	0.904	0.732	0.723	0.947	0.838	0.915	0.838	0.704	0.677	0.618	0.715	0.673	0.680	0.786
SegDiff	0.885	0.872	0.891	0.703	0.654	0.852	0.702	0.874	0.819	0.715	0.654	0.632	0.697	0.652	0.695	0.753
UNetr	0.926	0.936	0.918	0.785	0.702	0.969	0.788	0.893	0.828	0.732	0.717	0.554	0.658	0.683	0.722	0.762
Swin-UNetr	0.959	0.960	0.949	0.894	0.827	**0.979**	0.899	0.944	0.899	0.828	0.791	0.745	0.817	0.875	0.841	0.880
nnUNet	0.965	0.959	0.951	0.889	0.820	0.980	0.890	0.948	0.901	0.821	0.785	0.739	0.806	0.869	0.839	0.878
MedSegDiff	0.963	**0.965**	0.953	0.917	0.846	0.971	0.906	0.952	0.918	0.854	0.803	0.751	**0.819**	0.868	0.855	0.889
SAM 1 point	0.632	0.759	0.770	0.616	0.382	0.577	0.508	0.720	0.621	0.317	0.085	0.196	0.339	0.542	0.453	0.493
SAM 3 points	0.733	0.784	0.786	0.683	0.448	0.658	0.577	0.758	0.625	0.343	0.129	0.240	0.325	0.631	0.493	0.542
SAM 10 points	0.857	0.855	0.857	0.800	0.643	0.811	0.749	0.842	0.677	0.538	0.405	0.516	0.480	0.789	0.637	0.699
MedSAM 1 point	0.671	0.803	0.825	0.687	0.541	0.712	0.671	0.785	0.703	0.607	0.531	0.588	0.729	0.814	0.833	0.700
MSA 1-point	**0.968**	0.961	**0.959**	**0.926**	**0.861**	0.971	**0.919**	**0.960**	**0.928**	**0.863**	**0.825**	**0.767**	0.803	**0.879**	**0.862**	**0.893**

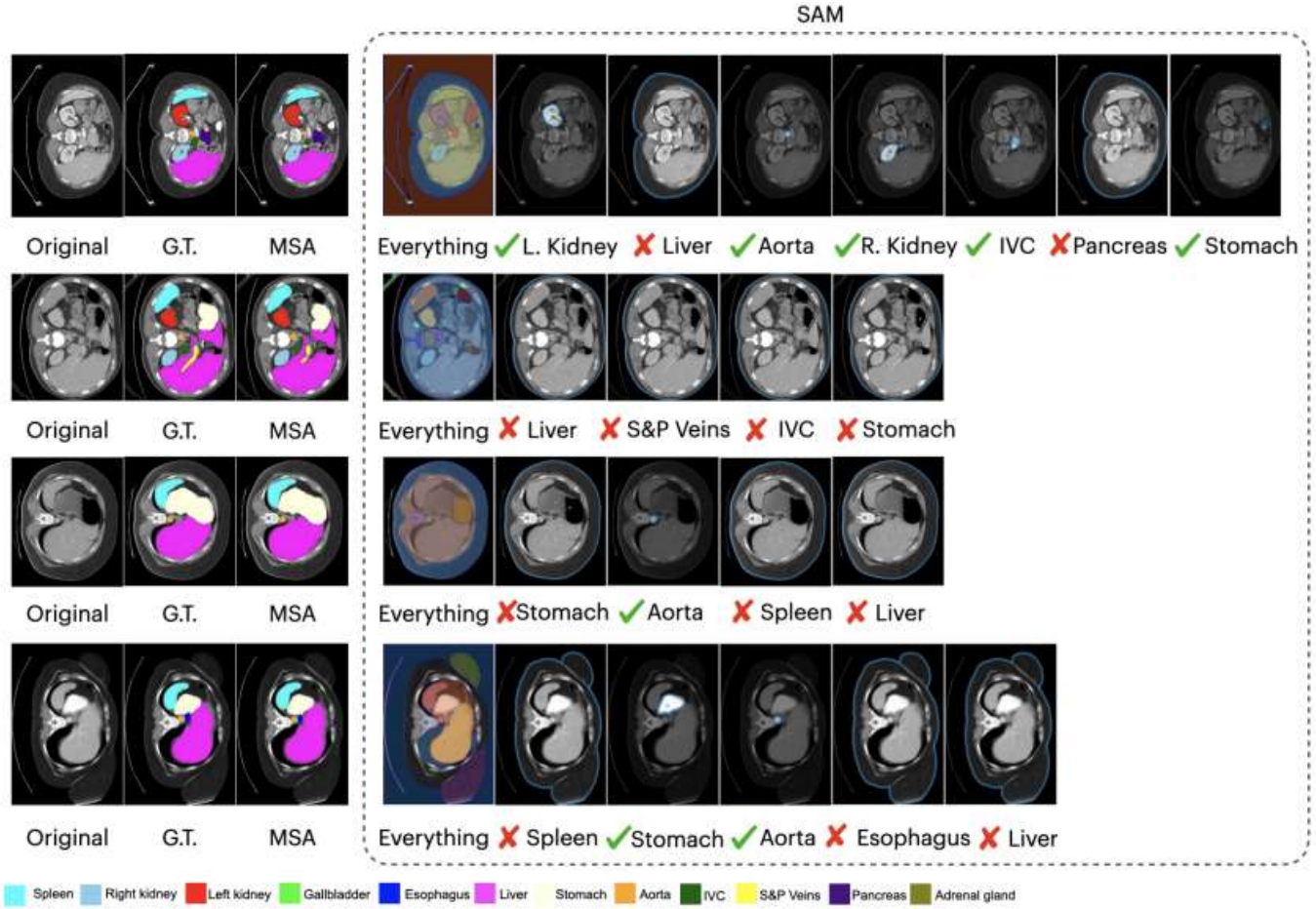
- SOTA model: The best current model in the research task
- SOTA result: The result of the best current model in the research task.

The comparison of MedSegDiff-V2 with SOTA segmentation methodsover BTCV dataset evaluated by Dice Score.

Model	Spleen	R.Kid	L.Kid	Gall.	Eso.	Liver	Stom.	Aorta	IVC	Veins	Panc.	AG	Ave
TransUNet	0.952	0.927	0.929	0.662	0.757	0.969	0.889	0.920	0.833	0.791	0.775	0.637	0.838
EnsDiff	0.938	0.931	0.924	0.772	0.771	0.967	0.910	0.869	0.851	0.802	0.771	0.745	0.854
SegDiff	0.954	0.932	0.926	0.738	0.763	0.953	0.927	0.846	0.833	0.796	0.782	0.723	0.847
UNetr	0.968	0.924	0.941	0.750	0.766	0.971	0.913	0.890	0.847	0.788	0.767	0.741	0.856
Swin-UNetr	0.971	0.936	0.943	0.794	0.773	0.975	0.921	0.892	0.853	0.812	0.794	0.765	0.869
nnUNet	0.942	0.894	0.910	0.704	0.723	0.948	0.824	0.877	0.782	0.720	0.680	0.616	0.802
MedSegDiff	0.973	0.930	0.955	0.812	0.815	0.973	0.924	0.907	0.868	0.825	**0.788**	0.779	0.879
SAM 1 points	0.518	0.686	0.791	0.543	0.584	0.461	0.562	0.612	0.402	0.553	0.511	0.354	0.548
SAM 3 points	0.622	0.710	0.812	0.614	0.605	0.513	0.673	0.645	0.483	0.628	0.564	0.395	0.631
SAM 10 points	0.785	0.774	0.863	0.658	0.673	0.785	0.760	0.712	0.562	0.703	0.651	0.528	0.704
MedSAM 1 point	0.751	0.814	0.885	0.766	0.821	0.901	0.855	0.872	0.746	0.771	0.760	0.705	0.803
MSA 1 point	**0.978**	**0.935**	**0.966**	**0.823**	**0.818**	**0.981**	**0.931**	**0.915**	**0.877**	**0.811**	0.767	**0.809**	**0.883**

- SAM's zero-sample performance in targeted medical image segmentation tasks is generally inferior to fully trained models
- MSA has the best overall performance

Visual comparison of MSA and SAM on abdominal multi-organ segmentation



use check mark to represent SAM correctly found the organ and cross to represent it lost.

- MSA accurately segments parts that are difficult for the human eye to recognize.

Visual comparison of MSA and SAM on abdominal multi-organ segmentation

	Optic-Cup		Brain-Turmor			Thyroid Nodule	
	Dice	IoU	Dice	IoU	HD95	Dice	IoU
ResUnet	80.1	72.3	78.4	71.3	18.71	78.3	70.7
BEAL	83.5	74.1	78.8	71.7	18.53	78.6	71.6
TransBTS	85.4	75.7	87.6	78.44	12.44	83.8	75.5
EnsemDiff	84.2	74.4	88.7	80.9	10.85	83.9	75.3
MTSeg	82.3	73.1	82.2	74.5	15.74	82.3	75.2
UltraUNet	83.1	73.78	84.5	76.3	14.03	84.5	76.2
SegDiff	82.5	71.9	85.7	77.0	14.31	81.9	74.8
nnUNet	84.9	75.1	88.5	80.6	11.20	84.2	76.2
TransUNet	85.6	75.9	86.6	79.0	13.74	83.5	75.1
UNetr	83.2	73.3	87.3	80.6	12.81	81.7	73.5
Swin-UNetr	84.3	74.5	88.4	81.8	11.36	83.5	74.8
MedsegDiff	85.9	76.2	**88.9**	**81.2**	**10.41**	84.8	76.4
SAM 1 points	-	-	63.2	58.6	25.53	-	-
SAM 3 points	-	-	65.5	61.7	24.87	-	-
MSA	**86.8**	**78.8**	87.6	81.2	12.46	**86.3**	**78.7**

The grey background denotes the methods are proposed for that/these particular tasks. Performance is omitted (-) if the algorithm fails over 70% of the samples

- MSA outperforms most models in segmentation

4.Conclusion

- Extension of **SAM segmentation model** to medical image segmentation named **MSA**.
- By employing **parameter-efficient adaptation**, achieved significant improvement over the original SAM model.
- Obtained state-of-the-art performance on **19 medical image segmentation tasks** across **5 different image modalities**.

Thank You!