SAM-Path: A Segment Anything Model for Semantic Segmentation in Digital Pathology

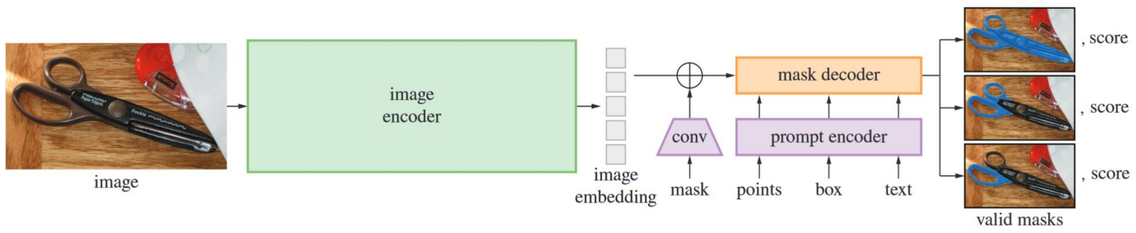
Jingwei Zhang[1], Ke Ma[2], Saarthak Kapse[1], Joel Saltz[1], Maria Vakalopoulou[3], Prateek Prasanna[1], Dimitris Samaras[1]
[1]Stony Brook University; [2]Snap Inc.; [3]CentraleSupélec, University of Paris-Saclay

MICCAI 2023 Vancouver CANADA MedAI

P1

Introduction

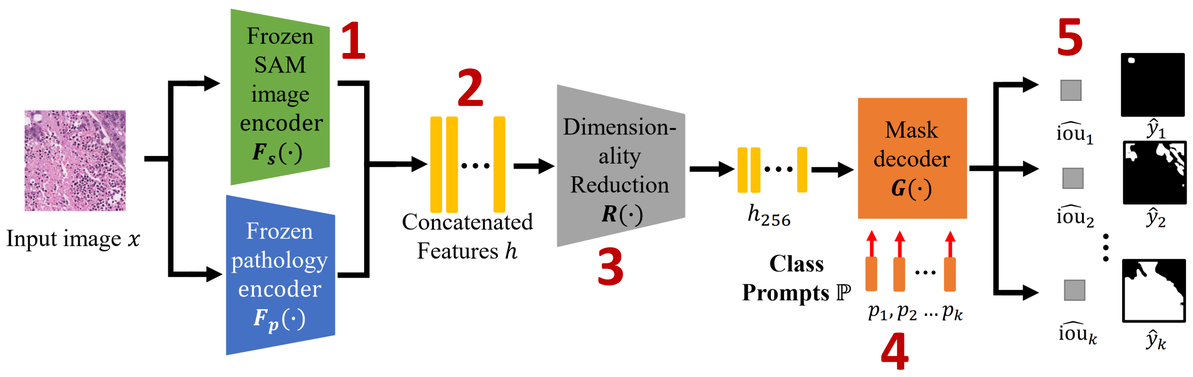
- Semantic segmentations is very important in computational pathology workflows [1].
- Segment Anything Model (SAM) [2], as a foundation models, has been recently proposed for universal segmentations.



- However, SAM **does not work** for semantic segmentation in pathology
  - We make it possible using **trainable class prompts** and an **additional pathology encoder**
- SAM has some three major limitations:
  - Lacking diverse **pathology** images in SAM training [3]
    ->suboptimal performance
  - Requiring **manually** inputted prompts
  - Lacking **semantic** predictions
- ***Our contributions:***
- ***SAM-Path*: Adapting** the vanilla SAM for semantic segmentation tasks using trainable prompts
- **Introducing** a **pathology foundation** model as an additional pathology encoder to provide **domain-specific** information

Method

- **Overview:**
  - Removing the prompt decoder of SAM
  - Adding a **pathology encoder**
  - Adding trainable **class prompt** P NEW!



1. Fed the input image into **frozen** vanilla **SAM** encoder and a **frozen pathology** encoder. We use HIPT [4] as the pathology encoder.

3. Concatenate the output features of batches.

2. Forward features into the **dimensionality reduction** layer to adjust the dimensionality of output features.

4. Input trainable **class prompts** to the mask decoder. If we have k classes, we should have k class prompts.

5. The mask decoder **predicts** segmentation masks and IOUs of each class.

- **Loss function:**

Loss weights (hyper-parameters)

$$\mathcal{L} = \sum_{i=1}^{k}[(1-\alpha)\mathcal{L}_{dice}(\hat{y}_i, y_i) + \alpha\mathcal{L}_{focal}(\hat{y}_i, y_i) + \beta\mathcal{L}_{mse}(\hat{iou}_i, IOU(\hat{y}_i, y_i))]$$

Dice loss | Focal loss | MSE loss

Two loss weights are optimized using grid search.

Conclusion

- SAM-Path introduces **trainable prompts** to SAM to identify classes.
- SAM-Path integrates a **pathology encoder** to incorporate more **domain-specific** knowledge.
- Our method **facilitates** semantic segmentation **without** the need for **manual** prompts.
- We plan to explore **generic SAM** in digital pathology in future work

Experiments

- **Datasets:**

BCSS [5]: Breast cancer, 20k annotations, 40X magnification
5 classes: Tumor, Stroma, Inflammatory, Necrosis and others

CRAG [6]: Colorectal cancer, 213 images, 20X magnification
1 class: gland

- **Comparing to vanilla SAM** on CRAG**:**

Methods	Dice	IOU
Vanilla SAM*	0.5245	0.3555
Vanilla SAM w. post-processing**	0.6598	0.4924
SAM-Path (ours)	**0.8841**	**0.8831**

36-53% higher | 22-39% higher

* We manually provide a dot annotation to the SAM.
* We assume all the segmented instances are glands.
** Filters out instances occupying more than half of the image as some instances occupy the entire image.

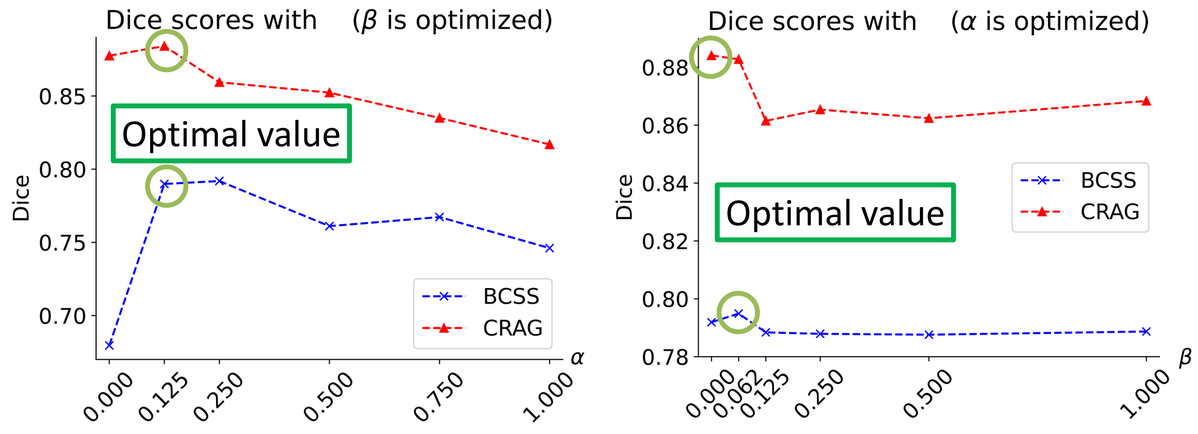
- **Comparing to fine-tuned SAM:**

Methods	BCSS	CRAG
Fine-tuned SAM w.o. pathology encoder	0.7562	0.8414
SAM-Path w.o. SAM encoder	0.7813	0.8191
SAM-Path (ours)	**0.7949**	**0.8841**

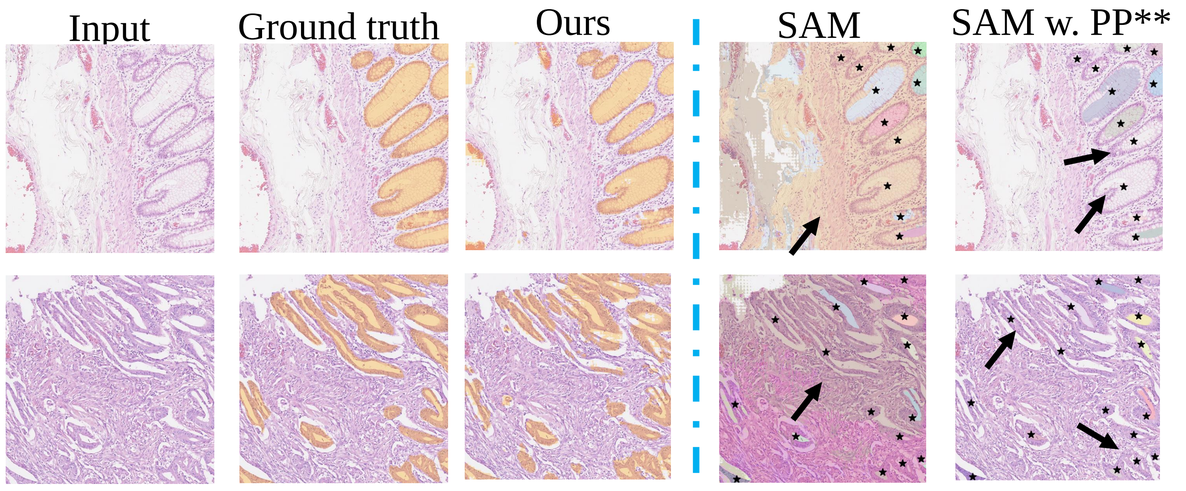
4% higher | 1-7% higher

The **pathology** encoder and vanilla **SAM** encoder are both necessary

- **Ablation study on two loss weights:**



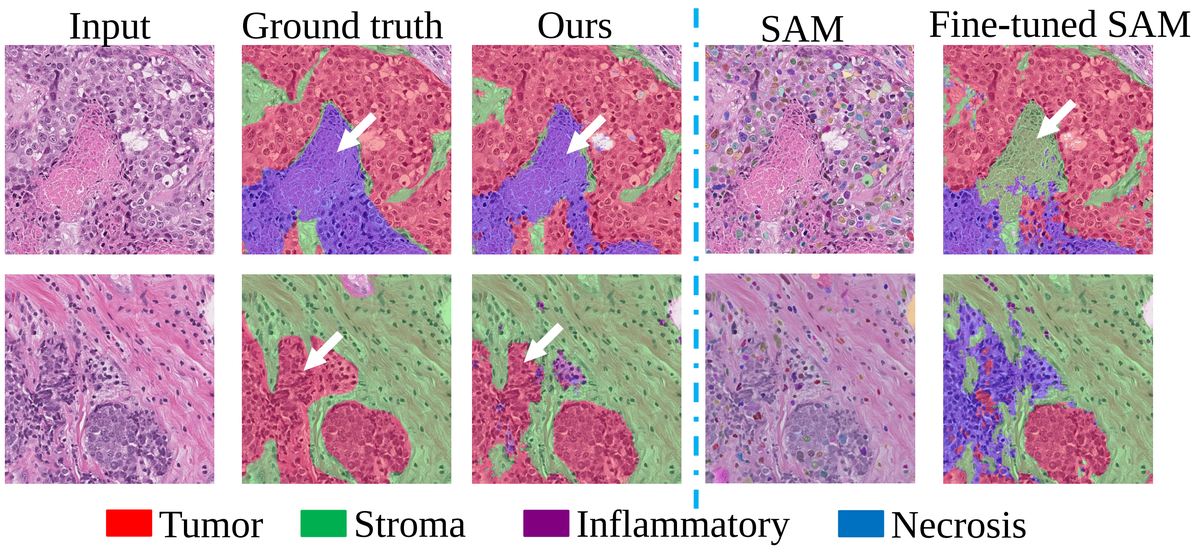
- **Qualitative analysis on CRAG*:**



* Black asterisks represent manual dot prompts we provide to the SAM.
** PP represents post-processing.

SAM-Path predicts better masks

- **Qualitative analysis on BCSS**



SAM-Path knows semantics and predicts better masks

[1] Litjens et al. A survey on deep learning in medical image analysis. In Medical image analysis (2017)
[2] Kirillov et al. Segment anything. In arXiv preprint (2023)
[3] Deng et al. Segment anything model (sam) for digital pathology: Assess zero-shot segmentation on whole slide imaging. In arXiv preprint (2023)
[4] Chen et al. Scaling Vision Transformers to Gigapixel Images via Hierarchical Self-Supervised Learning. In CVPR (2022)
[5] Amgad et al. Structured crowdsourcing enables convolutional segmentation of histology images. In Bioinformatics (2019)
[6] Graham et al. Mild-net: Minimal information loss dilated network for gland instance segmentation in colon histology images. In Medical image analysis (2019)

Acknowledgements: This work was partially supported by the ANR Hagnodice ANR-21-CE45-0007, the NSF IIS-2212046, the NSF IIS-2123920, the NIH 1R21CA258493-01A1, the NCI UH3CA225021 and Stony Brook University Provost Funds.