DH602

SAM + CLIP Based Universal Model for Multi-Organ Segmentation

GROUP 17

Manav Ketan Doshi
Mohit Kedia
Pratik Shah
Shreedhar Malpani

Project Update

Brief Recap

01 MODELS

02 DATASETS

03 RELATED WORK

04 PROPOSED ARCHITECTURE

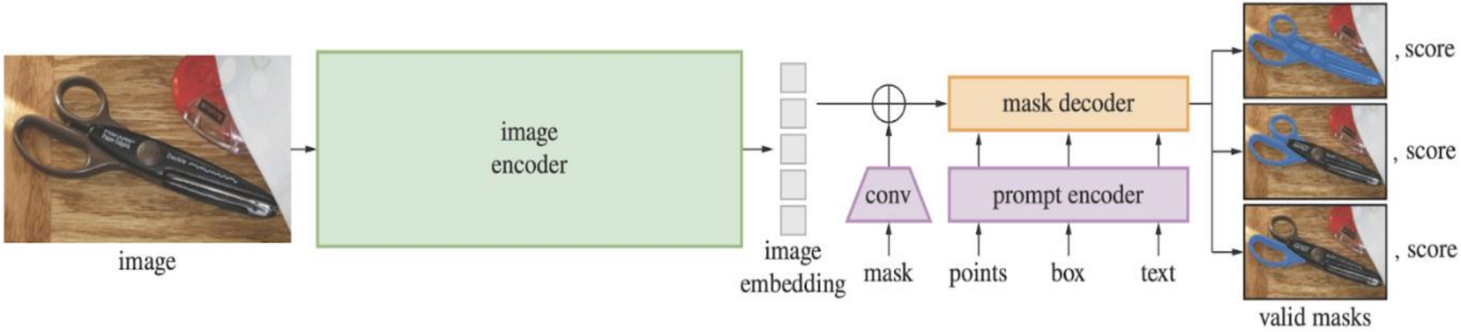
Aim

Our main objective is to utilise the power of recent large scale models such as CLIP and SAM to create an universal model that can segment multiple organs along with predicting them (their class) from a given input image.

01

Models

SAM Model



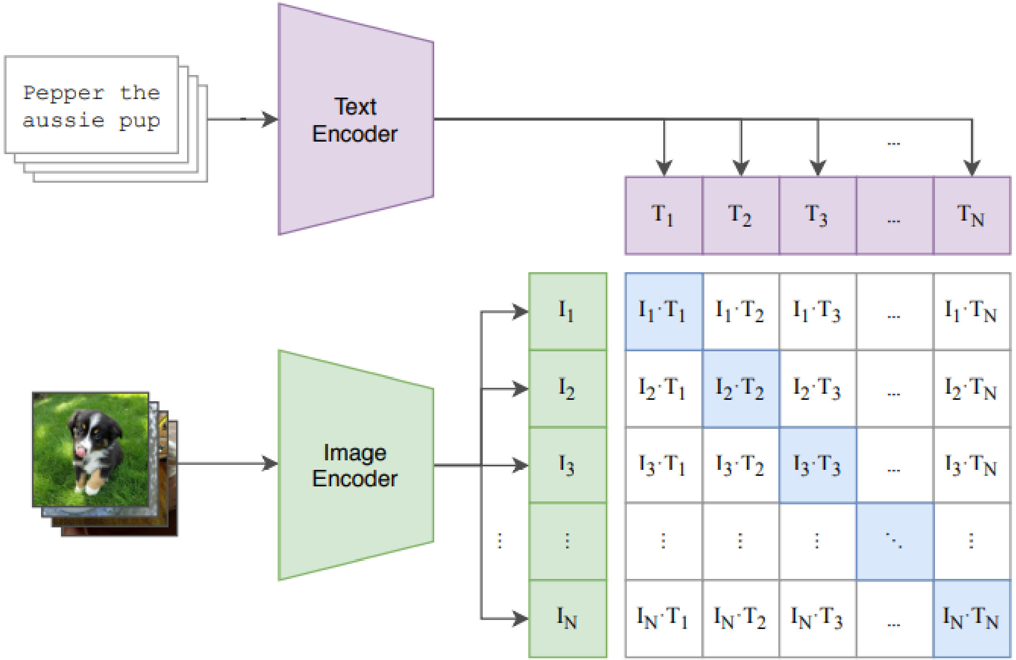
The model can be used to predict segmentation masks of any object of interest given an input image. SAM is designed for a promptable segmentation task with a prompt in the form of a point, bounding box, or text. The model tries to predict a segmentation mask for the region indicated by the input prompt.

CLIP Model (1/2)

In the last few years, pre-training methods which learn directly from raw text have revolutionized NLP. This suggests that information grounded in natural language can be used for pretraining image-based models more efficiently. While such approaches have been used in the past, they have been difficult to scale up due to the enormous corpus of text data available.

Given N pairs of images and captions, CLIP is trained to predict which of the pairings across a batch actually occurred. To do this, CLIP learns a multi-modal embedding space by jointly training an image encoder and text encoder to maximize the cosine similarity of the image and text embeddings of the N real pairs in the batch while minimizing the cosine similarity of the embeddings of the N^2 - N incorrect pairings.

CLIP Model (2/2)



Pepper the aussie pup

Text Encoder

Image Encoder

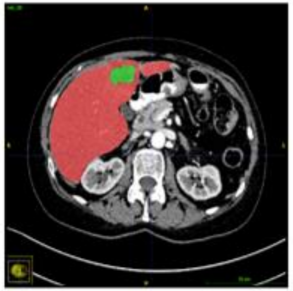
	T_1	T_2	T_3	...	T_N
I_1	$I_1 \cdot T_1$	$I_1 \cdot T_2$	$I_1 \cdot T_3$	...	$I_1 \cdot T_N$
I_2	$I_2 \cdot T_1$	$I_2 \cdot T_2$	$I_2 \cdot T_3$	...	$I_2 \cdot T_N$
I_3	$I_3 \cdot T_1$	$I_3 \cdot T_2$	$I_3 \cdot T_3$	...	$I_3 \cdot T_N$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
I_N	$I_N \cdot T_1$	$I_N \cdot T_2$	$I_N \cdot T_3$	...	$I_N \cdot T_N$

02

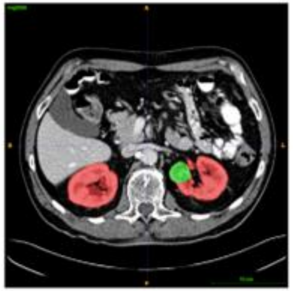
Datasets

Datasets utilized

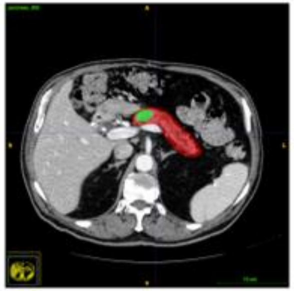
- We are going to use 6 public datasets of CT scans with different annotated body organs for training.
- There are a total of 6 annotated major body organs in the datasets- Pancreas, Liver, Kidney, Spleen, Lung and Bladder
- We also have public datasets with 19 more minor body organs, which we can use after success with 6 organs.



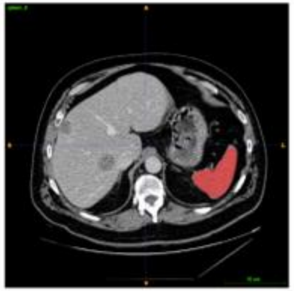
Liver



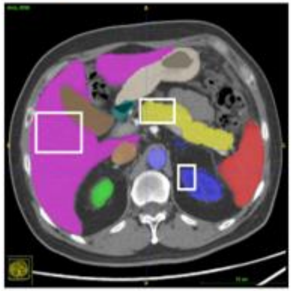
Kidney (L & R)



Pancreas



Spleen



Ideal (To segment major organs)

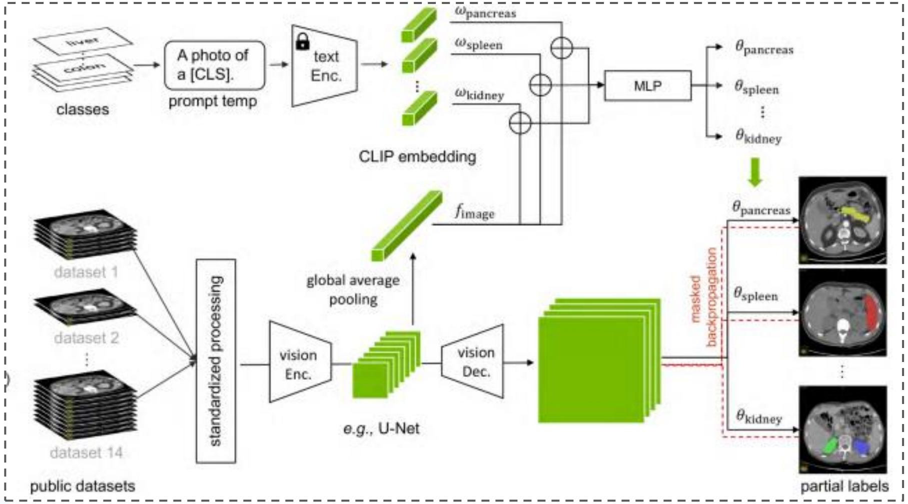
	Dataset	#Targets	#Scans	Annotated Organs
Training	Pancreas-CT	1	82	Pancreas
	LiTS	1	201	Liver
	KiTS	1	300	Kidney
	AbdomenCT-1K	4	1112	Spleen, Kidney Liver & Pancreas
	CT-ORG	4	140	Lung, Liver, Kidneys & Bladder
	CHAOS	3	40	Liver, Kidneys & Spleen
Testing	TotalSegmentator	6	1128	All

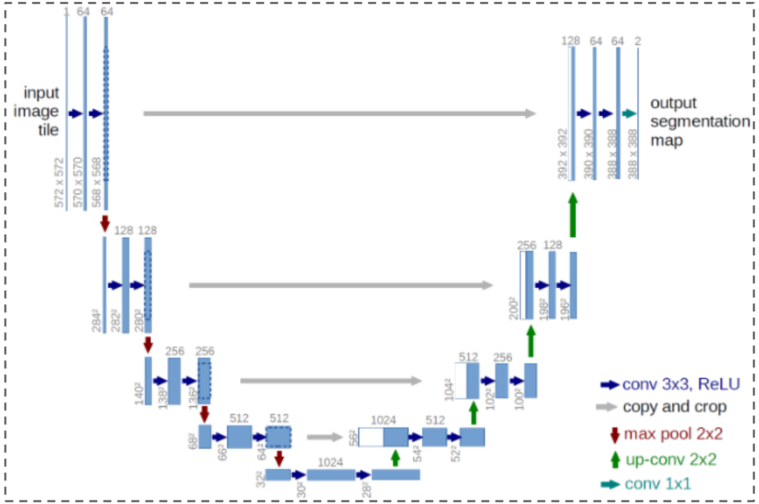
03

Related Work

CLIP-driven Universal Model for Organ Segmentation and Tumour Detection

A similar architecture was used in this paper and instead of using SAM they have used U-Net for image segmentation. Here is the architecture -





Reference: https://arxiv.org/pdf/2301.00785.pdf

Segment Anything in Medical Images

SAM is found to be useful in various medical applications, like liver tumour segmentation, brain MRI segmentation, CT organ segmentation, surgical instrument segmentation, and more.

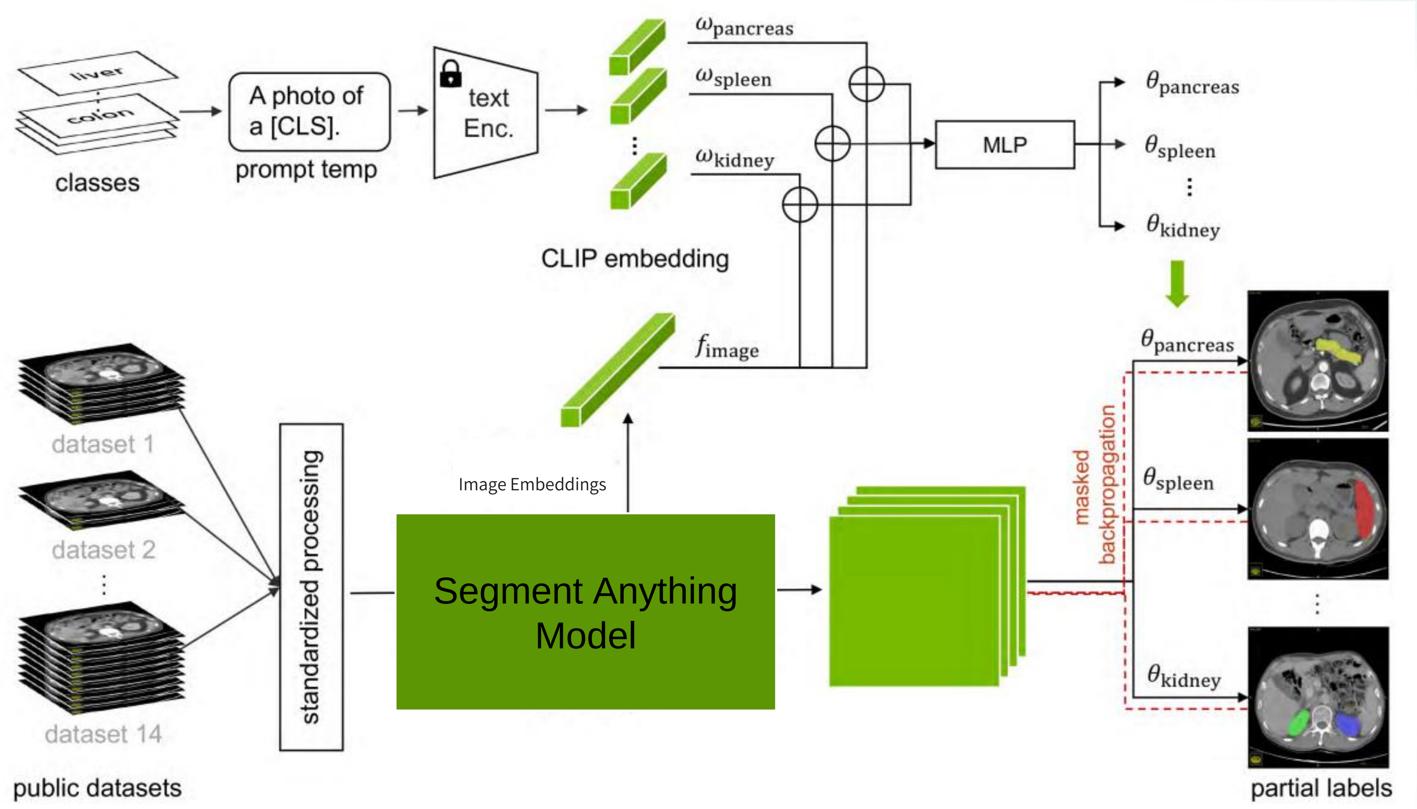
In cases of whole slide images (WSI), when checked for tumour tissue segmentation and cell segmentation, it is found that SAM accurately segments large objects

SAM's performance improves significantly when provided with good prompts. Two main prompt types are points and boxes.

04

Proposed Architecture

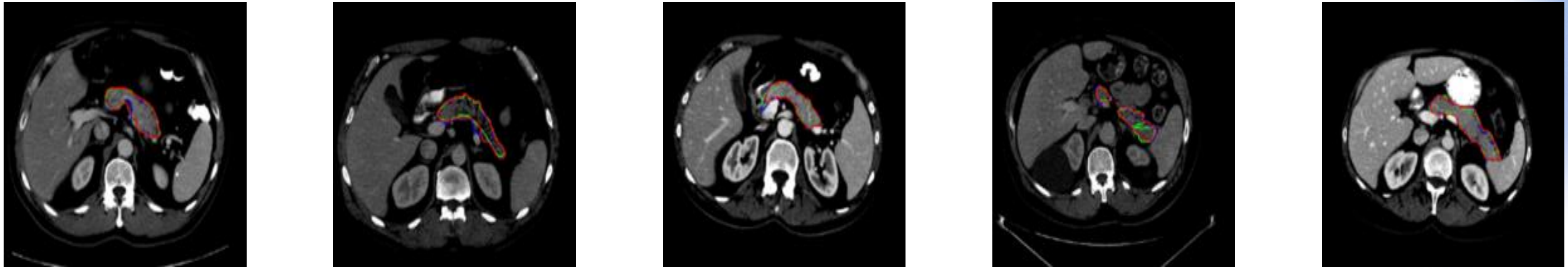
Proposed Architecture



Proposed Architecture

- We pass the same prompt to the CLIP and SAM models.
- The image embeddings from the SAM model are then concatenated with CLIP embeddings and then passed through the MLP network.
- Let F denote the output of the SAM model. The prediction for each class is computed as $\boldsymbol{P}_k = \text{Sigmoid}(((\boldsymbol{F} * \boldsymbol{\theta}_{k_1}) * \boldsymbol{\theta}_{k_2}) * \boldsymbol{\theta}_{k_3})$
- Here the theta values are outputs of the MLP network
- We can also try fine tuning the SAM model on the input labelled images that we have

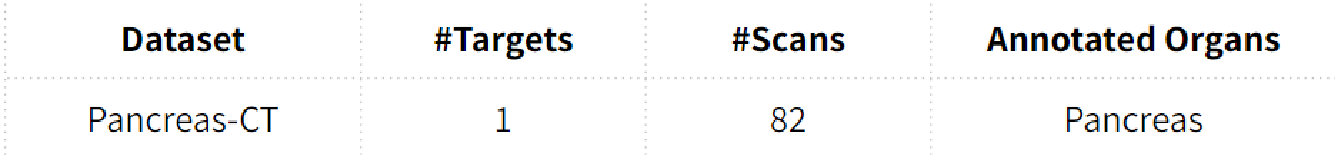
Results of the original Architecture



Represents the result from the original model where the red line represents the annotation from Doctor 1; green line indicates the annotation from Doctor 2; blue line shows the results generated by Universal Model.

Implementation of the Original Architecture

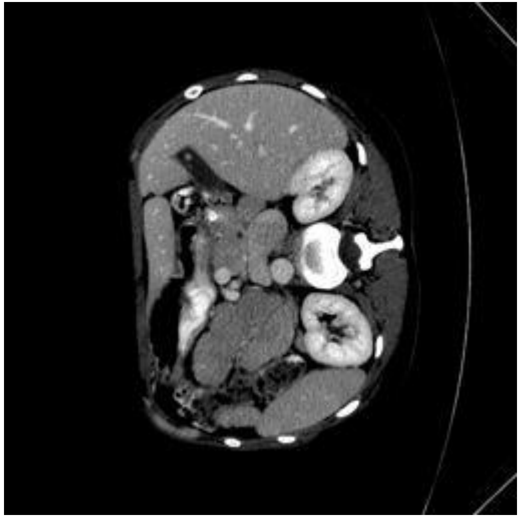
- Right now we are testing the model on the first dataset -



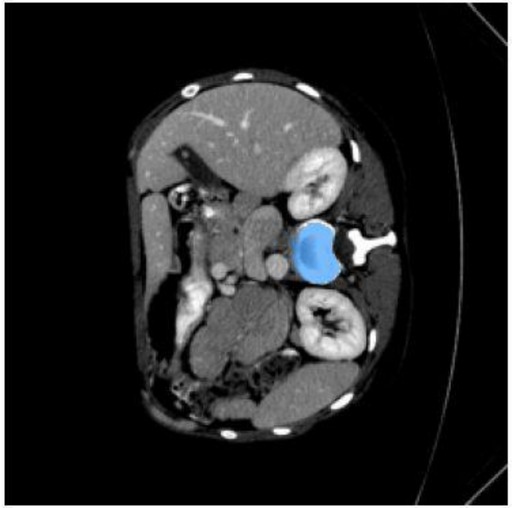
Dataset	#Targets	#Scans	Annotated Organs
Pancreas-CT	1	82	Pancreas

To see how well the model responds to the single label. Then we will move on to combining more datasets for achieving our outlined goal of multi-organ segmentation.

Implementation of the Original Architecture



Predicted mask



References

- **Github Repo: https://github.com/shreedharmalpani/DH-602-2024/upload/main**
- **CLIP-Driven Universal Model for Organ Segmentation and Tumor Detection(https://arxiv.org/pdf/2301.00785.pdf)**
- **CLIP (https://arxiv.org/abs/2103.00020)**
- **SAM (https://arxiv.org/abs/2304.02643)**
- **Liver Tumour Segmentation-** Chuanfei Hu and Xinde Li. When sam meets medical images: An investigation of segment anything model (sam) on multi-phase liver tumor segmentation. arXiv preprint arXiv:2304.08506, 2023
- **Brain MRI Segmentation -** Mohapatra, A Gosai, and G Schlaug. Sam vs bet: A comparative study for brain extraction and segmentation of magnetic resonance images using deep learning. arXiv preprint arXiv:2304.04738, 2:4, 2023
- **CT Organ Segmentation** - Saikat Roy, Tassilo Wald, Gregor Koehler, Maximilian R Rokuss, Nico Disch, Julius Holzschuh, David Zimmerer, and Klaus H Maier-Hein. Sam. md: Zero-shot medical image segmentation capabilities of the segment anything model. arXiv preprint arXiv:2304.05396, 2023
- **Surgical Instrument Segmentation** - An Wang, Mobarakol Islam, Mengya Xu, Yang Zhang, and Hongliang Ren. Sam meets robotic surgery: An empirical study in robustness perspective. arXiv preprint arXiv:2304.14674, 2023
- **Tumour Tissue Segmentation and Cell Segmentation** - Ruining Deng, Can Cui, Quan Liu, Tianyuan Yao, Lucas W Remedios, Shunxing Bao, Bennett A Landman, Lee E Wheless, Lori A Coburn, Keith T Wilson, et al. Segment anything model (sam) for digital pathology: Assess zeroshot segmentation on whole slide imaging. arXiv preprint arXiv:2304.04155, 2023