MedSAM and UNet: A Comparative Analysis for Surgical Tool Segmentation

Lola Assad - 20266807, Jackson Reid - 20267039, Adam Cockell - 20158798

Motivation

Image segmentation is important to the medical domain. Its use cases range from diagnosis with radiological scans, disease monitoring, tumor monitoring, artifact detection, anomaly detection, tissue identification, pathological cancer screening, cell growth tracking, and onwards with more uses every year with the development of methods in artificial intelligence. Surgical tool segmentation specifically is to be used in real-time surgical interventions for improved accuracy of care and the integration of robotics in surgical procedures.

Problem Description

Using a dataset of real surgical gastrointestinal images where DaVinci tools are present in the set of images, our team conducted a comparative analysis of UNet and Med-SAM for separating the surgical tools from organic tissues. The team chose Med-SAM as one of our comparative models for a variety of reasons. Med-SAM is open source making it accessible. There is also sufficient documentation, a tutorial, and a GitHub repository from the Bo Wang laboratory on its usage and development. It is a fine-tuned version of the Segment Anything Model for the medical domain making it well-suited as a pseudo-control for our purposes. And finally, there is an option to incorporate a bounding box through an interactive graphical user interface.

Med-SAM was designed by the Bo Wang lab at the University of Toronto by fine-tuning the Segment Anything Model from MetaAI. The Segment Anything Model is composed of an image encoder, whose embeddings are then sent to a corresponding decoder where masks are produced. The difference between this model and Med-SAM is that the Med-SAM architecture involves the addition of a fine-tuned head. The data run through this fine-tuned head include computed tomography scans, magnetic resonance images, surgical scene images, surgical scene images with tools like the DaVinci 6 robot, and pathological data including dyed slides for cellular-level analysis. This large breadth of medical data on top of a pre-existing segmentation model makes med sam well suited for a large array of tasks in the medical imaging domain.

We chose UNet as a backdrop for the Med-SAM comparison because of UNets suitability for its feature extraction capabilities, and unique network architecture. UNet is widely used in biomedical image segmentation largely due to its extensive convolutional layers, pooling, and outputted accurate masks. Due to the course's time constraints, the objective was to ensure the pixel-wise separability of the DaVinci robot and internal organs.

Contribution

Lola Assad - Med-SAM Analysis
Jackson Reid - Binary mask creation and started the U-Net model's code.
Adam Cockell - Model programming and tuning

Related Work

Study: "Segment Anything in Medical Images"

This study investigates the performance of Med-SAM across a diverse range of medical segmentation tasks. The goal was to build a single model that could segment a variety of medical images from different modalities such as X-rays, MRIs, and CT scans. The accuracy of this model was compared to state-of-the-art medical image segmentation models. In comparison, Med-SAM performed inconsistently, with significant variation in accuracy depending upon the specific imaging context. Generally, Med-SAM performed exceptionally when faced with images containing distinct borders, but struggled with images containing less defined boundaries.

The researchers involved in this study created a dataset of 1,570,263 image-mask pairs. The data includes 10 modalities and many different types of cancer.

Study: "Medical Image Segmentation Using Automatic Optimized U-Net Architecture Based on Genetic Algorithm"

This study aimed to improve U-Net performance with medical image segmentation tasks by updating and optimizing its architecture using a genetic algorithm. To do so, the algorithm explores various U-Net configurations and selects the optimal design for a certain segmentation task. These configurations were assessed using the Dice similarity coefficient and the Intersection over Union metrics to measure accuracy. The newly optimized architecture achieved higher accuracy than the baseline model, showing how effective genetic algorithms can be in improving model performance. A great benefit of this model is the reduction in the necessity for manual tuning. On the other hand, however, this adds serious complexity to the model.

This study used three datasets, all used to segment parts of the images. The first was a lung image dataset used to segment the lung. The second was a cell nuclei dataset used to segment the nuclei. The last was a liver dataset used to segment the liver regions.

Study: "Half-UNet: A Simplified U-Net Architecture for Medical Image Segmentation"

The goal of this study was to reduce the complexity of the U-Net architecture, aiming to decrease the number of parameters and the computational demands, while maintaining or improving performance. To do this, the Half-Unet model simplifies the encoder and decoder components of U-Net by unifying the channel number, implementing full-scale feature fusion, and incorporating ghost modules to streamline the network. Using these methods, Half-UNet achieved segmentation accuracy similar to the original U-Net and some of its variants. Regarding its goal of reduction, it decreased parameters by 98.8% and floating-point operation by 81.6% compared to the standard U-Net. This helped with the model's complexity, but greatly hindered the generalization potential of U-Net, demanding the model undergo extra validation than normally required.

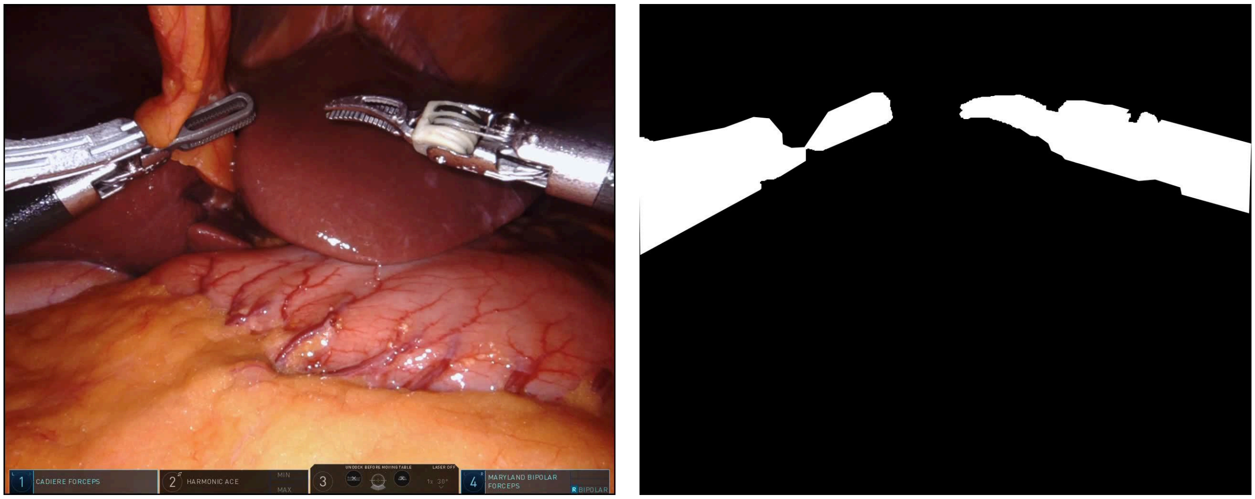
This study used three medical image datasets. The first was a mammography dataset used to segment breast tissue abnormalities. The second was a lung nodule CT image dataset that was used to detect potentially malignant nodules in lung tissue. The last was a dataset of left ventricular MRIs used to test how the model performed in delineating cardiac structures.

Dataset

Our dataset includes images from three cross-validation sets of real image data from 40 surgical videos of gastrectomy for gastric cancer Additionally, the dataset includes many computer-generated images of the same gastrectomy surgery. The images include 14 different surgical instruments and six organs. An issue that came with so many instruments is that some of them had similar colors to tissue in the images, making classification more difficult. Another challenge was the presence of dark and blurry regions along the edges of many images. These visual inconsistencies complicated the process of accurately creating masks for the images, resulting in lower model performance and accuracy.

The dataset included masks for each image, where the masks were made so that each medical instrument could be segmented into several parts. This went against our original plan for the study, which was to classify parts of the image as tissue or tool. To overcome this challenge, we manually created masks for the U-Net implementation to pursue the original idea. A big downfall of this method is that the mask creation was done by hand which introduced human error into the model.

Sample real image and mask:



Individual Implementation

Lola Assad

I. Data Preprocessing

To preprocess the image data, I first loaded them from a common file, then resized all images using OpenCV to have (3, 256, 256) dimensionality.

II. Experimental Set-Up

For this step, I used a function to generate a mask using the MedSAM model to separate the DaVinci tools from the organic tissue.

III. Training, Testing, Validation

MedSAM is a pre-trained model with a fine-tuned head composed of medical images ranging from radiological images, cell and tissue pathology dyed slides, and other various medical image types. For the testing, I ran the real surgical images through the model after preprocessing and then analyzed how frequently a correct separation was made to calculate the accuracy.

Jackson Reid / Adam Cockell

I. Data Preprocessing

We extracted a subset of 33 images from the real image dataset folder and created binary masks for each. The images are resized to 128x128 and then normalized before being used to create the model.

II. Experimental Set-Up

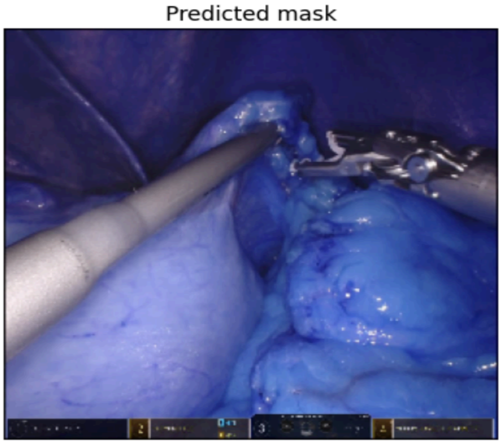
Gimp was used to manually create masks separating the classes in the image. The model is built to distinguish between 2 classes - the tools and the tissue - thus the masks are two-tone black and white.

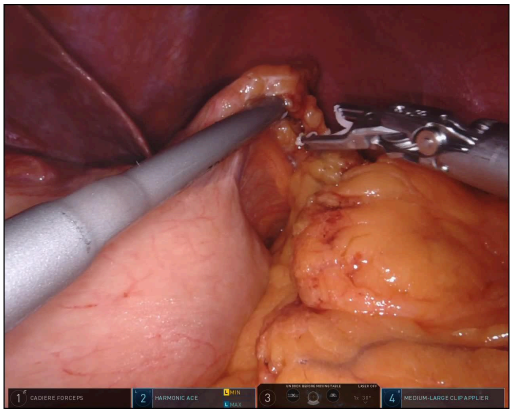
III. Training Testing Validation

25 images (75%) are used to train the model, with testing being done on the remaining 8 images (25%) for 20 epochs (iterations). The Adam optimizer was used along with the binary cross-entropy loss function, and accuracy and mean IoU metrics for fitness. These were used as they are widely accepted as the best-performing or most balanced options for the task of image segmentation in a U-Net. Once the training is complete, a window containing the real image, real mask, and predicted mask for each image in the testing group is visually displayed for inspection and validation. Various tracked metrics are also printed to console over each epoch.

Results and Discussions

Med-SAM Results - Lola





In around 97% of the cases we tested, Med-SAM was able to produce a precise pixel-wise segmentation of the surgical tools and organic tissue (accuracy: 97%). Med-SAM was trained on 1,570,263 image-mask pairs, covering 10 imaging modalities and over 30 cancer types. It was tested on the wide medical image dataset. For metrics of evaluation, the dice similarity coefficient was used with normalized surface distance. For statistical analysis of Med-SAM, the Wilcoxon signed-rank test was used to compare paired samples.

UNet Results

The generated model achieves an accuracy of approximately 88%, with loss values around 30 and mean IoU of 0.42. Visually, the predicted masks resemble the real masks fairly accurately with accuracy loss mainly arising from some areas of the DaVinci tools not being detected.

Model	Accuracy
MedSAM	97%
UNet	88%

This level of accuracy was achieved with only 20 samples, with significant gains possible given a larger dataset. Due to the difficulty in producing accurate masks and additional computational requirements, we decided against increasing the number of images in the dataset. U-Nets also tend to have rapidly diminishing returns in accuracy relative to dataset size, and we believe this sample is sufficient to demonstrate the potential of the model.

Conclusion and Future Work

Our team can conclude that while our U-Net is capable of surgical segmentation, the wide array of training data that went into the development of Med-SAM, along with the SAM by MetaAI architecture with a fine tuned medical imaging head, made Med-SAM more suited to the task of surgical tool segmentation. The accuracy of the Med-SAM model beat the accuracy of UNet.

In future iterations of our model, we could work towards video segmentation, and three-dimensional surgical scene reconstruction using both UNet and Med-SAM. This has applications in robotic-assisted surgical interventions and could improve the accuracy and precision of care, improve surgeons' visual bandwidth in procedures, and improve the efficiency of care. It is also important to make these models accessible for use in lower-income populations, our team would work towards the development of a model requiring less computational power, without sacrificing accuracy.

References

1. Ma, Jun, et al. "Segment anything in medical images." Nature Communications 15.1 (2024): 654.
2. Ma, Jun, et al. “Segment Anything in Medical Images.” Nature Communications, vol. 15, no. 1, 22 Jan. 2024, arxiv.org/pdf/2304.12306v2.pdf, https://doi.org/10.1038/s41467-024-44824-z.
3. “MedSAM.” GitHub, 2 Dec. 2023, github.com/bowang-lab/MedSAM.
4. Yoon, Jihun. “Surgical Scene Segmentation in Robotic Gastrectomy.” Kaggle.com, 2022, www.kaggle.com/datasets/yjh4374/sisvse-dataset, https://doi.org/10.1007/978-3-031-16449-1_53). Accessed 5 Dec. 2024.