Machine Learning Project Report

Shengfan Wang(2200012978),Siyuan Song(2200012935),Yang He(2000013166)

January 15, 2024

Abstract

The main focus of this project is to apply the SAM model to the BTCV dataset. The objective is to assess the performance of the model on the given dataset, and accordingly fine-tune its mask decoder. Additionally, the aim is to modify the model's structure to enable it to generate label outputs with classification capabilities. Moreover, the project involves contemplating scenarios for other potential applications.

1 Task 1

In Task One, we utilized SAM's official pre-trained model to perform organ segmentation on two-dimensional slices of the test data. We compared the segmentation results using various prompts (single point, multiple points, bounding box). In the single-point mode, we randomly selected a point from the objects in the image as the prompt. For the multiple-point mode, we chose multiple random points within each object as prompts, considering cases where there might be an insufficient number of points, in which we included all available points. Finally, in the bounding box mode, we attempted to select the minimum bounding box surrounding each object as the prompt. The results are presented in Table 1.

For the code implementation,our approach is to first use the label to identify all possible grayscale values, which represent different organs. Next, we still utilize the label to select prompts for each organ (single point, multiple points, bounding box). Finally, we train using the official training model of SAM. After training, we use the generated mask to calculate the mean Dice coefficient and output the results.

In the process of implementation, figuring out how to calculate the mean Dice coefficient took up a lot of our time. We initially thought that the mean Dice was related to the score, but after reviewing the literature, we didn't find any relevant records. It wasn't until Professor Wang mentioned in class that the network would also output labels that we realized we needed to calculate the mean Dice coefficient ourselves.Thus,in the final step, within the array of Dice coefficients, we identified the complete set of Dice coefficients for each organ and calculated the mean Dice coefficient (mDice) for that organ. Then, we computed the mDice for all organs.

From the data in the table, it's not hard to see that the bounding box approach yields the best results, significantly outperforming the others, while the effectiveness of the remaining methods is relatively similar.

Prompt Type	Mean Dice
single point	0.582
three points	0.605
seven points	0.604
bounding box	0.751

Table 1: Mean Dice between BTCV and output

2 Task 2

In Task Two, we fine-tuned the mask decoder for a model designed to segment arbitrary objects. The optimization was based on using a single point as the prompt. We utilized the Adam optimizer and adopted $(1 - meanDice)^2$ as the loss function. After the fine-tuning process, the performance of the model surpassed that of the base model.

We first initialize and configure a deep learning model for image segmentation tasks, setting up the Adam optimizer with a learning rate of 10^{-5}. We perform 10 iterations per execution, using backpropagation to calculate the gradients. Finally, the model results are saved in a file named '`module_vit_h.pth`'.

Additionally, we experimented with other loss functions, such as $loss = 1 - meandice$ and $loss = sigmoid(\frac{1}{score})$. We also tried adjusting the learning rate to 10^{-4}. These experiments correspond to the files '`try_all_vit_h.pth`', '`try1_vit_h.pth`', and'`try2_vit_h.pth`', respectively.However, none of these performed as well as '`module_vit_h.pth`'

It is evident that when we further train and fine-tune based on the results obtained using a single point as the prompt, the results show a minimal improvement.

Prompt Type	Mean Dice	New Mean Dice
single point	0.582	0.586
three points	0.605	0.607
seven points	0.604	0.606
bounding box	0.751	0.753

Table 2: Difference

3 Task 3

In Task Three, we modified SAM by adding a neural network for classification after segmentation. This enhancement enabled the model to output labels for each detected object. To achieve this, we integrated a CNN structure into the system. Subsequently, we began training the model with a dataset. The final accuracy achieved 0.3,and allows for effective identification of organs(Figure 1 and 2 display a successful recognition.).

This CNN network is structured as follows:The first section consists of a two-dimensional convolutional layer with 32 3x3 convolution kernels and a 'relu' activation function, followed by a pooling layer. The second section includes a convolutional layer with 64 convolution kernels and a max pooling layer. The third section comprises a convolutional layer with 128 convolution kernels and a max pooling layer. The fourth section features a flattening layer that transforms the output of the previous convolutional layers into a one-dimensional array, preparing it for input into a fully connected layer. The fifth section contains a fully connected layer with 128 neurons, continuing to use the 'relu' activation function. The sixth section is also a fully connected layer, equipped with 48 neurons, and is used for classification.Since we are unable to determine which organs correspond to the different grayscale values, the CNN we designed recognizes all the organs that have appeared, including the thirteen types of organs required by the project.

Previously, we attempted to preprocess images using SAM with pre-trained parameters and used the feature vectors from the intermediate results(we think the return parameter of the transformer in the mask decoder,hs, which SAM use to predict the iou of the results,has the ability to predict the classes for organs with MLP network)as input for an MLP to perform category prediction. However, after trying this, we found that it was worse than that of a CNN, and we couldn't identify the cause of the issue.

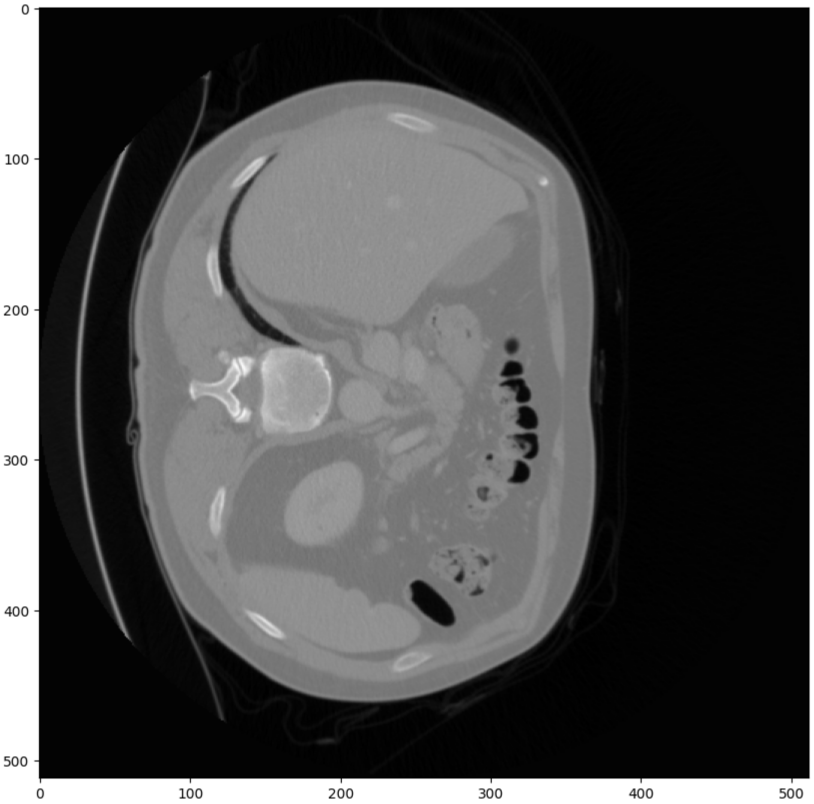
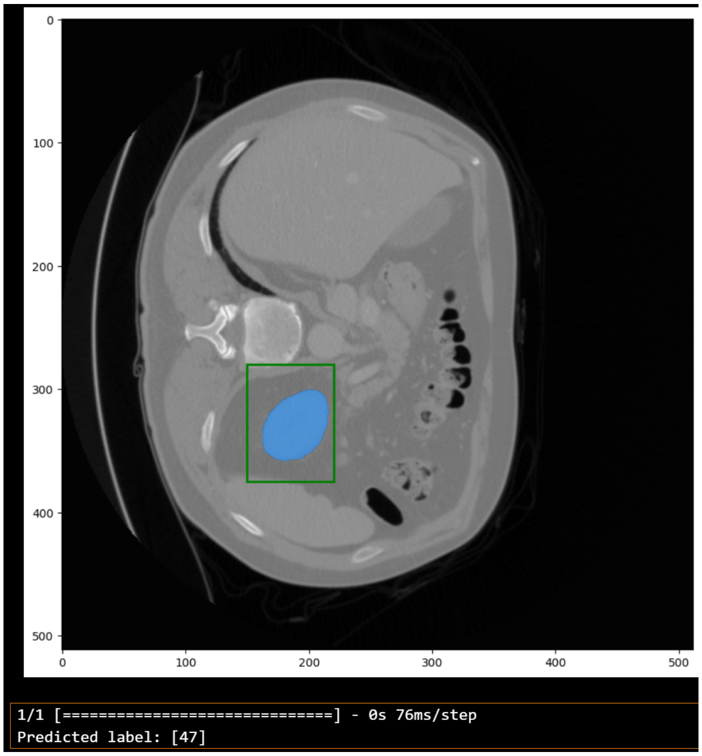


Figure 1:Unrecognized image



```
1/1 [==============================] - 0s 76ms/step
Predicted label: [47]
```

Figure 2:Recognized image

4 Task 4

Periodontitis is one of the most common chronic infammatory disease afecting half of the adults in the USA. It is initiated by bacterial bioflm infection of the periodontal soft and hard tissues, including the gingiva, the cementum, the periodontal ligament, and the alveolar bone. We can use the SAM model to do the RBL distance.We can use SAM to do segmentation for each tooth in order to make us easy to get the important points for the calculation of the RBL distance.We can use SAM combining with the traditional computer vision method such as Harris corner,line detector and so on.I think the SAM will make the estimated RBL distance more accurate.

In the medical field, in order to meet the needs of disease diagnosis and treatment plan formulation, it is often necessary to scan patients to determine the condition of various internal organs. Before the emergence of deep learning methods, this process was mainly completed directly by doctors. After the emergence of deep learning, we can save doctors' work and improve the accuracy of judgments. The first difficulty in brain region segmentation is to distinguish between the brain and non brain (such as the skull) regions. In the segmentation of MRI images, the brightness of brain tissue is a very important feature. However, due to the presence of noise, partial volume effect (PVE), bias field effect, and other factors in MRI images, brightness based segmentation algorithms are prone to misjudgment. In order to achieve relatively accurate segmentation, there are several commonly used MRI data preprocessing methods, and one important operation is background voxel removal. Its purpose is to extract brain tissue and separate it from non brain tissues (such as fat, skull, neck, etc.) that may have brightness overlap with brain regions, thereby assisting in internal segmentation of brain regions. In the process of segmenting the brain and non brain parts, we can use the SAM model to optimize the segmentation. Using a SAM to recognize the brain and non brain parts would be an optimization to improve recognition accuracy