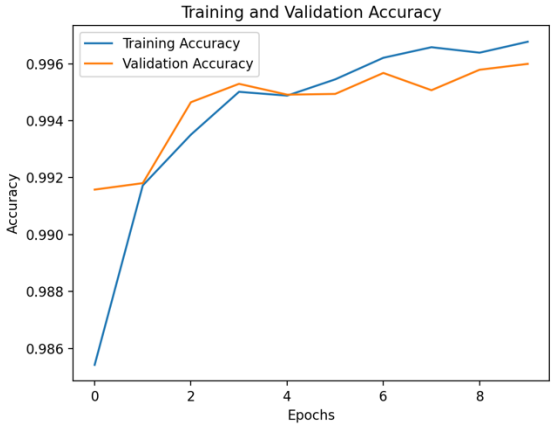
Ngoc Bao Han, Nguyen
20188794
CISC 472
April 11th, 2023

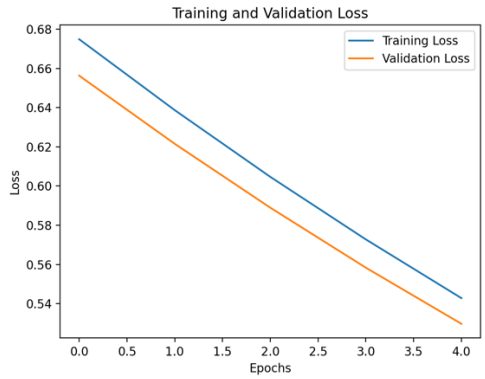
Assignment 4 report

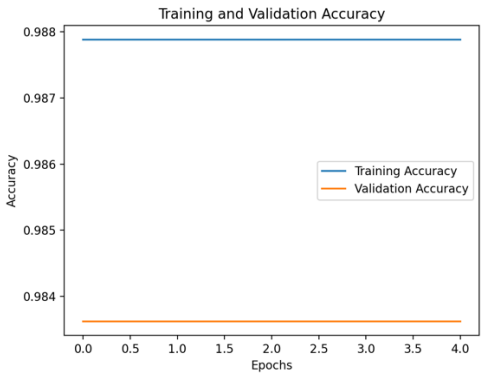
Question 1.

This is the two plots of training loss and training accuracy I've produced after running Unet.



I tried and run with a smaller learning rate (0.001 to 0.00001), but the run above was still better.



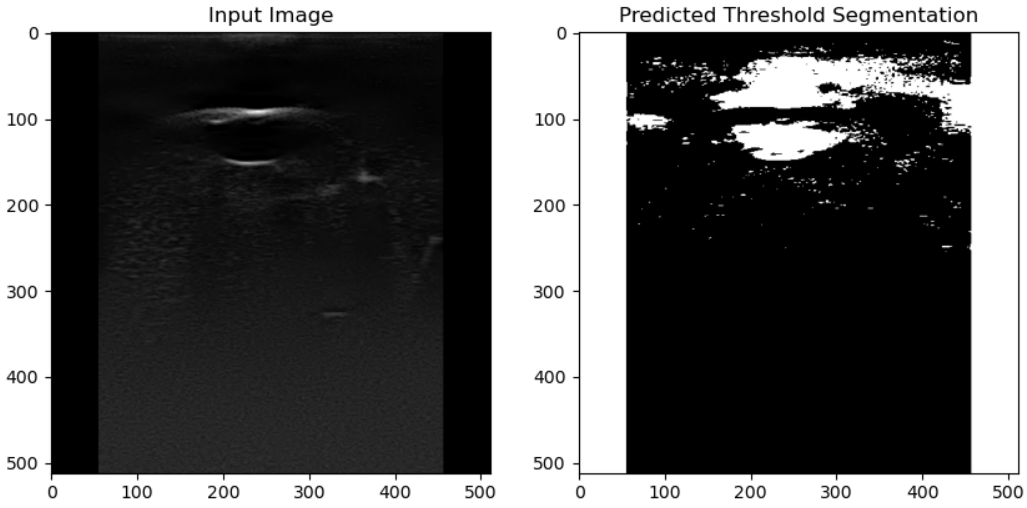


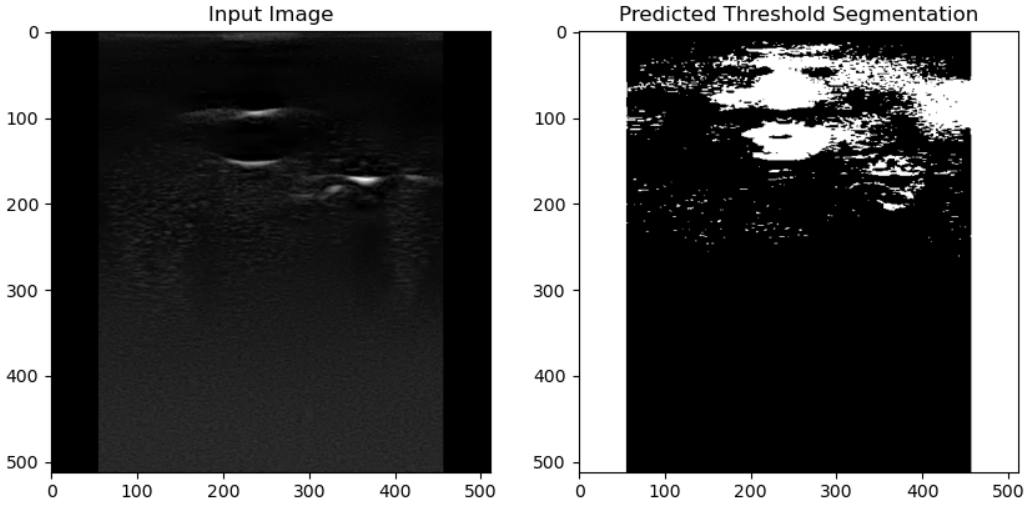
Question 2.

Threshold segmentation experimentations:

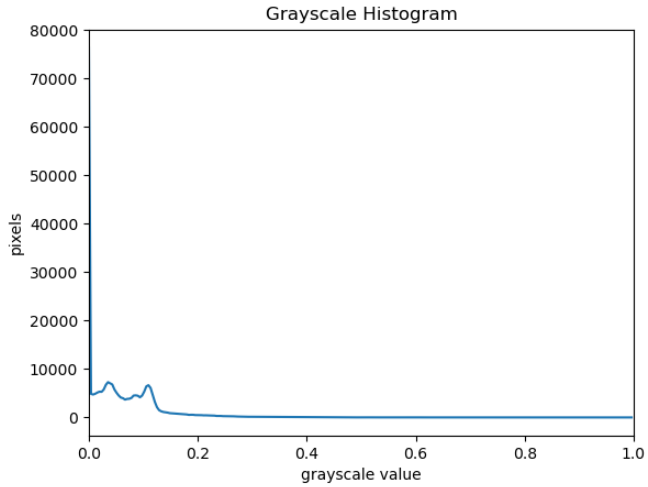
- Manual thresholding: The first thought that I have for choosing threshold values is to understand that are most common pixels value that the segmented area has. The vessel image in the ultrasound tends to have a darker value than its surrounding and the background pixels. The value of the segmented area tends to fall between [0, 4].
  The result I get with manual thresholding is quite good on test data. Though it is still considering some of the background as the vessel. This is a disadvantage of threshold

segmentation where pixel value is the indicator of object, we know that for this task, it's simply not that straightforward.

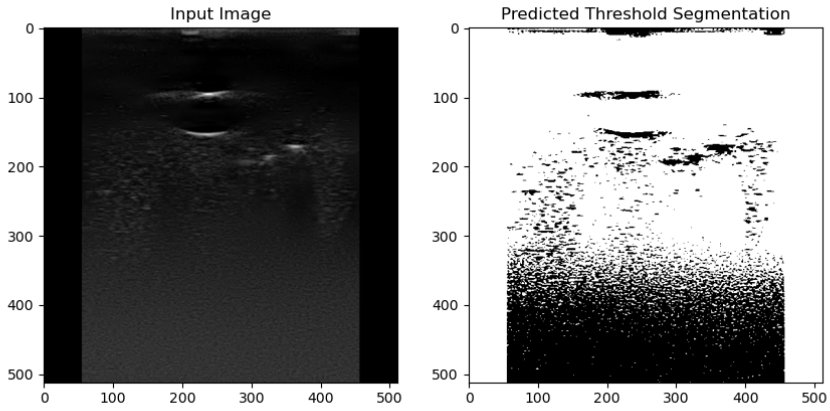


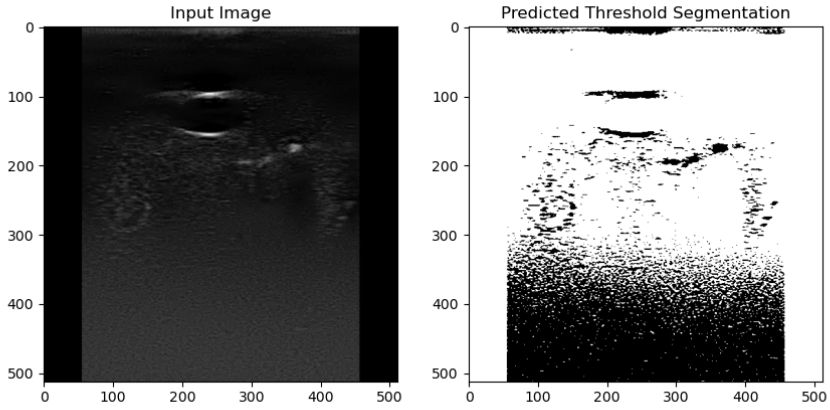


- Histogram based thresholding: This approach involved determining the threshold value based on the histogram of pixel intensity in the image. The basic idea is to choose a threshold value that separates the image into foreground and background regions. A common approach after the histogram has been generated is the Otsu's method. In the experimentation code for Question 2, I randomly pick 10 images out of the training set, create histograms of those and gather Otsu value from those histograms. All 10 of the histograms come out looking similar. I will attach along a histogram from the batch below. I went ahead and calculate the mean Otsu value for all the images from the training set and got value: 0.1986936613043667. This means that any pixel between 0 and 19 (Otsu value) is one class, and other pixels are another class.



A few images from thresholding with Otsu value.





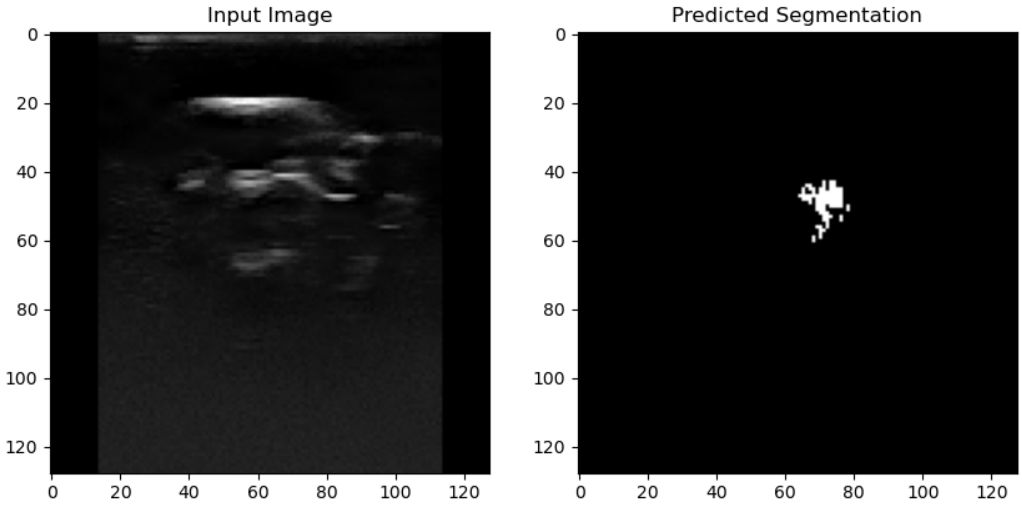
Region-growing segmentation experimentation:

- A little confession: My region-based segmentation function from assignment 2 is incorrect. The reason being is that I should only assign the first seed a value of 1, not

both. This means that my calculation for pixel intensity is incorrect and thus the potential segmentation created by this will be flawed as well. However, I tried to figure out a few ways to choose a proper seed values for region-growing.

UNet segmentation experimentation:

For me, there isn't much to experiment with the Unet segmentation, I went ahead and just produce segmentation images from the trained model.



Question 3.

N/A

Question 4.

N/A

Question 5.

N/A

Question 6.

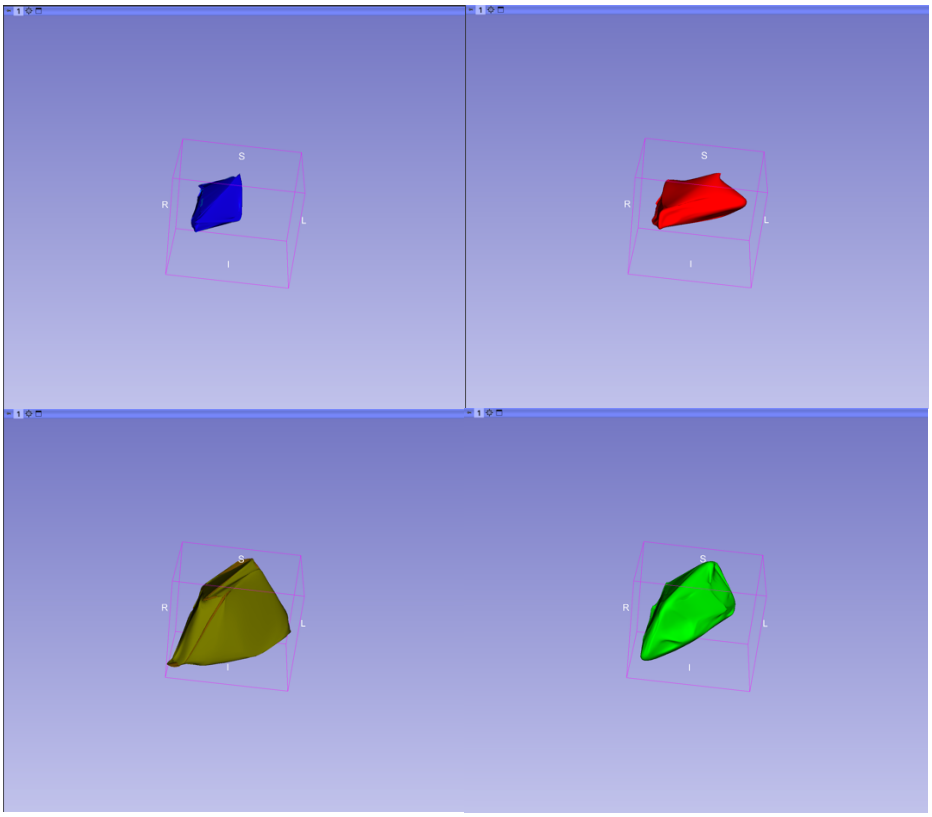
N/A

Question 7.

Qualitative evaluation of models:

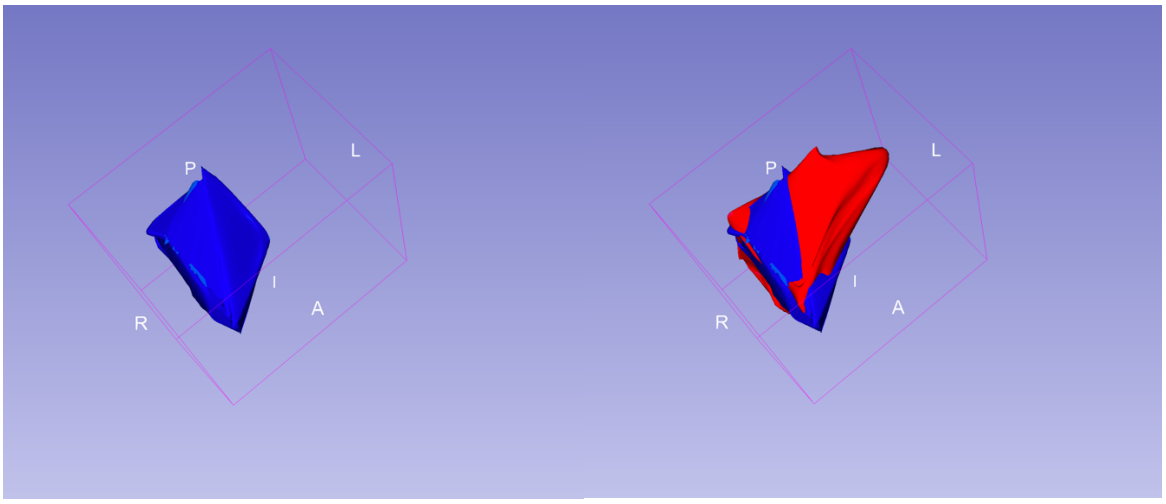
From left to right, top to bottom, the order of model shown are:

(1) Ground truth – blue.
(2) Unet segmentation – red.
(3) Threshold segmentation – brownish green.
(4) Region growing segmentation – neon green.



Observations:

- In terms of size in increasing order, the ground truth is the smallest, the Unet is the second smallest, following by region growing, and threshold segmentation is the largest.
- In terms of shape and resemblance, I would say Unet model has the most accurate segmentation, following by region growing, and threshold.
- The Unet segmentation captures the size and shape of the segmentation the closest, I have included a picture of the ground truth on its own below and an image of the ground truth model along with the Unet segmentation. We can see that the Unet is slightly larger. Some possible explanations that I can come up with are:
    - Over segmentation: This could be possible since the Unet model is trained on a large dataset (~2000 slices) that includes images with complex structure and contain outliers. Thus, could possibly result in over segmentation by detecting more objects or regions than are present in the ground truth.
    - Resolution mismatch: When training Unet, we had to rescale the images down from 512x512 to 128x128, this could results in the model trying to detect more structures or regions than the present in the ground truth.



- The threshold segmentation is the largest and has the lease texture overall. A possible speculation I have are:
    - Threshold selection: The threshold value used for segmentation might be too low, causing the segmentation to include more pixels than the ground truth. We have seen this in Question 2 where the threshold includes a lot more noise around the vessel than the actual vessel itself.
- The region growing segmentation is still quite larger than the ground truth, larger than the Unet segmentation but not as large as the threshold segmentation. A possible reason for this is:
    - Misalignment or noise in selecting seed and parameters: The region growing algorithm might be initialized with incorrect seed points and have suboptimal parameters such as the similarity threshold or the region size constraint. This could cause the algorithm to merge regions that should not be merged, leading to over-segmentation.

Question 8.

1. Volume and Surface Area calculation

	Surface Area	**Volume**
Groundtruth segmentation	6654.15mm^2	27206.43mm^3
Threshold segmentation	16763.71mm^2	119332.51mm^3
Region-growing segmentation	13752.74mm^2	111900.93mm^3
Unet segmentation	8507.48mm^2	46560.66mm^3

2. Average Accuracy and IoU
   Threshold IOU score 0.040780007368974434
   Region growing IOU score 0.07314680658083708
   Unet IOU score 0.12437270249121285

Threshold accuracy score 0.716016602406304
Region growing accuracy score 0.45774273498816426
Unet accuracy score 0.9156752731393559

Some conclusions:

The IOU and accuracy scores calculated suggest that the segmentation methods did not perform very well in reconstructing the vessel, as all the scores are quite low. Specifically, the Unet method had the highest IOU score of 0.124, indicating that it had the best overlap with the ground truth segmentation, but it still only captures 12.4% of the total area of the ground truth segmentation.

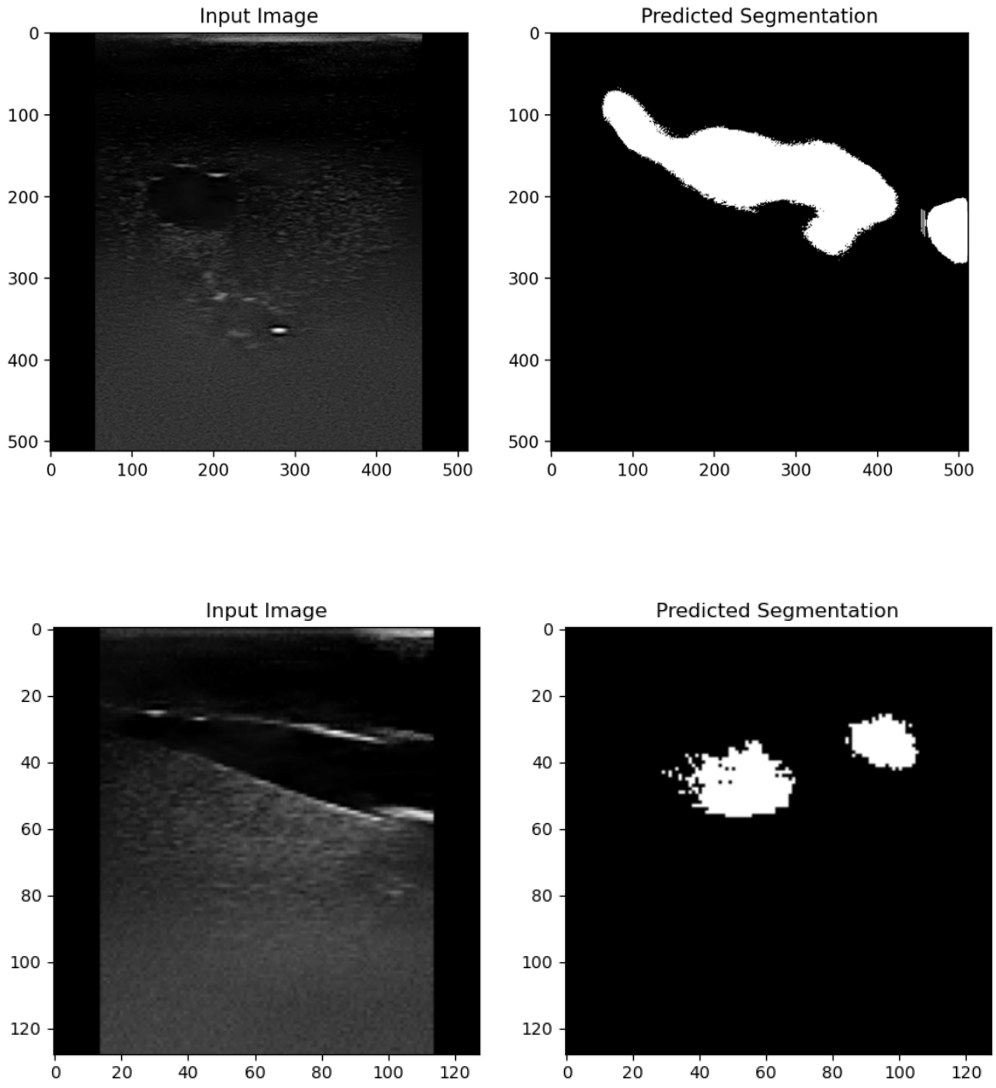
When considering the surface area and volume estimates, all segmentation methods produced larger volume estimates compared to the ground truth segmentation, with the threshold method producing the largest volume estimate of 119332.51mm^3. This suggests that all the segmentation methods tended to over segment the vessel, resulting in larger volume estimates. The same can be said after examining surface area data.

Overall, these results suggest that there is significant room for improvement in the segmentation methods used to reconstruct the vessel. While the Unet method had the highest IOU score, it still only captures a small portion of the ground truth segmentation, suggesting that the segmentation method may have been too conservative. The threshold method had the highest accuracy score, but it produced the largest volume and surface area estimates, suggesting that it may have been too aggressive in segmenting in the vessel. The region growing method performed somewhere in between the other two methods.

Question 9.

I start to run out of time here so I wasn't able to run any more experiments. However, if I were to have more time, these are potential tests that I'd run to improve accuracy and reconstruction on Unet segmentation since that was my top running model.

- Training the model without resizing the image to 128 x 128: The reason for over segmentation is low image resolution. Increasing the size of input images can help capture more details and reduce the likelihood of the model over generalize and over predict pixels when segmenting the vessel. I did get to run this part, but after training for 5+ hour my computer crashes so I could get to build a reconstruction and lost all my new segmented file. But the top if a segmention with the 512x512 quality, the bottom is from 128x128 quality.



If I have more time I would run this again, save all the segmentations and create a model from the newly segmented file on higher quality

- Tweaking the learning rates for the Unet model so loss and accuracy converges slower. It turns out that the predictions get worse in quality, the segmentation are patchy and there isn't any cohesive segments in the slices to work with
- Training data augmentation: Creating random rotations, flips, zooms, and translations. This will generate more variations to the training set and allow the model to avoid overfitting.
- Increasing model depth: Adding more layers to Unet model to help capture more complex features, this will work hand in hand with training data augmentation to avoid overfitting.
- Ensemble learning: Combining multiple Unet models to improve the accuracy of the final predictions This can be done by training several models with different

initializations or different data subsets, and then averaging their predictions at test time.

- Using a pretrained encoder: Such as VGG or ResNet, to improve the feature extraction capabilities of the model, especially when the training data is limited.