Final Project Report for CS 184A/284A, Fall 2023

Project Title: Segment Breast Ultrasound Images

List of Team Members:
Sohyun Shin, 71686312, sohyuns1@uci.edu

Yujeong Yang, 39208092, yujeony@uci.edu

Xinyu Zhang, 81641551, xzhang52@uci.edu

1. Introduction and Problem Statement

Breast cancer remains a significant health concern, being the most common cancer among women in the United States and the second leading cause of cancer-related deaths [1]. Early detection is pivotal in improving outcomes, and it is recommended to start screening every year for average women 45 or older [2]. This project aims to enhance breast cancer diagnosis through advanced technologies applied to breast ultrasound images. Specifically, we focus on detecting, and segmenting abnormal lesions within these images. The study utilized data from 600 patients, aged between 25 and 75 years, and employed a Faster R-CNN model as a baseline. We compared this baseline model with segmentation models specialized in medical images; U-Net, a convolutional neural network (CNN) tailored for medical images, and MedSAM, a vision transformer trained on medical data. Evaluation metrics used were precision and IoU to gauge segmentation quality. By doing so, the project aims to answer pivotal questions: 1) What segmentation model works the best for the breast ultrasound images? 2) In the specific context of breast ultrasound images, does the transformer-based segmentation approach demonstrate superior accuracy and efficiency compared to other segmentation models? Successful segmentation of breast ultrasound images will lead to tracking development of legions in quantitatively, in terms of their size and shape.

2. Related Work on the Issue

The questions we ask align with the broader goal of refining breast cancer diagnosis by integrating cutting-edge technologies and methodologies, contributing to the growing body of knowledge in the field. As of December 2021, CNNs have been widely employed for their ability to extract valuable features from images [3]. While our project incorporates CNN methodologies, it distinguishes itself by adopting state-of-the-art segmentation models. Due to the novelty of these models, limited research has been conducted, with available studies like “BreastSAM: A Study of Segment Anything Model for Breast Tumor Detection in Ultrasound Images” [4], “Breast cancer detection from ultrasound images using attention U-nets model” [5], “Comparative Analysis of Segment Anything Model and U-Net for Breast Tumor Detection in Ultrasound and Mammography Images” [6], primarily addressing challenges associated with segmenting complex boundaries. Our study seeks to bridge this gap by assessing the performance of these advanced segmentation models in the context of breast ultrasound images, comparing them against established methods, and contributing valuable insights to the field.

3. Data Sets

The data [7] collected at baseline include breast ultrasound images among women in ages between 25 and 75 years old. This data was collected in 2018. The number of patients is 600 female patients. The dataset consists of 780 images with an average image size of 500*500 pixels. The images are in PNG format. The ground truth images are presented with original images. The images are categorized into three classes, which are normal, benign, and malignant. Below is the example of the original input image and the image with mask.

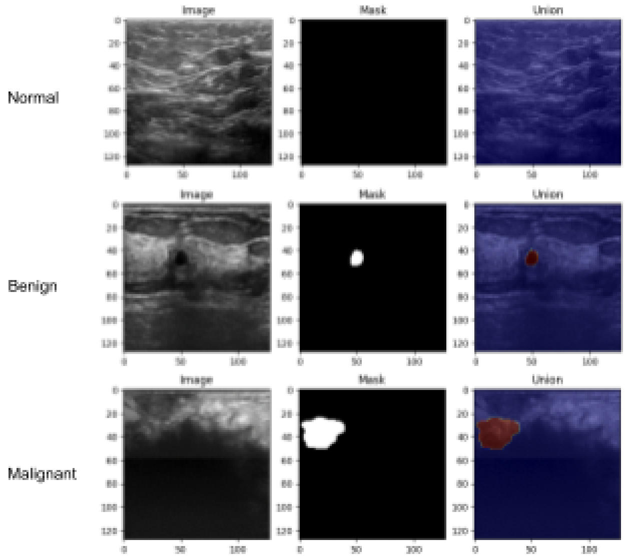


Figure 1: A comparative view of different classes of breast cancer ultrasound images, including the original scans, corresponding segmentation masks, and the overlaid images with masks highlighting the areas of interest.

4-1. Description of Technical Approach

In this project, we have implemented three different segmentation models in order to examine how well the segmentation models work in ultrasound images. We then compared these results with the regular detection model using faster R-CNN as a baseline model. The three segmentation models we utilized were Mask R-CNN, U-Net and MedSAM.

In the realm of segmentation, we opted for Mask R-CNN, serving as a benchmark against which other segmentation models were evaluated. This model extends Faster R-CNN by predicting segmentation masks on each Region of Interest (RoI), enabling precise object detection and pixel-level instance segmentation.

U-Net, recognized as a biomedical image segmentation algorithm, played a pivotal role. It assigned class labels to individual pixels and delineated object boundaries. Through convolution and symmetrical deconvolution paths, U-Net extracted feature maps, shedding light on the distribution of lesions within the image and their categorization.

MedSAM [8], a specialized medical image segmentation model derived from SAM (a vision transformer; ViT), addressed the unique challenges posed by medical images. Trained on an extensive dataset comprising over a million medical image-mask pairs, MedSAM exhibited proficiency in segmenting lesions and distinguishing between benign and malignant cases. Unlike traditional CNNs, which rely on convolutional layers to process image patches, ViT employs a transformer architecture and is already pretrained on large-scale datasets. This means that it is versatile and reduces the need for extensive training on task-specific datasets, but it is also computationally expensive since the model is trained on extensive and high-resolution images.

4-2 Software

For U-Net, we used the original model introduced by Ronneberger et al [10]. On the basis of the original structure, the model was then modified, implemented and evaluated by Yujeong Yang.

For MedSAM pretrained model developed by BoWang's Lab was used [8]. The code used to image preprocessing to fit the model's requirement, the fine-tuning of the model and also create and prompt bounding boxes were written by Sohyun Shin. For examining the results of MedSAM, Monai, a pytorch-based open source library for medical imaging, was used.

For Faster R-CNN model and Mask R-CNN model we use the Detectron2 framework. We utilized a small batch size of 2 images per batch due to hardware constraints, while setting the learning rate to 0.00025 and applying weight decay for regularization.

5. Experiments and Evaluation

5.1 Data preprocessing

In the experiment, we partitioned the dataset into three subsets: training, testing, and validation, with a ratio of 8:1:1, respectively. To enhance the model's ability to generalize, we employed a series of data augmentation techniques. These included resizing images to ensure the shortest edge was 800 pixels, applying random horizontal flips, and adjusting brightness, contrast, and saturation. Following the augmentation process, the class distribution within each subset of the dataset is presented below (Table 1).

5.2 Faster R-CNN

The model chosen for this task was a Faster R-CNN with a ResNet-50 backbone and Feature Pyramid Network (FPN). This model is well-suited for object detection tasks due to its balance between accuracy and computational efficiency.

The training process was monitored using three key loss metrics: total loss, classification loss (Loss CLS), and bounding box regression loss (Loss Box Reg). The total loss, representing the sum of all individual loss components during training, showed a significant decrease, indicative of the model's improving performance. The plot of classification loss exhibits a steady decline, with minor fluctuations, eventually plateauing, whereas the bounding box regression loss showing larger variations in loss values throughout the training. And it shows that whereas the model could classify the image into different categories, it faces challenges in precisely localizing objects in the context of ultrasound images.

5.3 Mask R-CNN

In this experiment, we employed the Mask R-CNN which extends Faster R-CNN architecture, to detect and classify breast ultrasound images into three categories. The model was configured using the Detectron2 framework. We utilized a small batch size of 2 images per batch due to hardware constraints, while setting the learning rate to 0.00025 and applying weight decay for regularization. Gradient clipping was enabled to stabilize training, which proceeded for 1000 iterations. The dataset was split into training (BUSI_train) and validation (BUSI_val) sets, processed by 8 data loader workers in parallel.

Figure 3 presents four different aspects of loss during the training process. The top left graph shows the total loss, which decreases steadily over 1000 iterations, indicating overall model convergence. The top right graph tracks the classification loss (Loss CLS), which also demonstrates a downward trend, reflecting the model's improving accuracy in classifying the input data. The bottom graph represents the regression loss (Loss Box Reg), which fluctuates but generally trends upwards, suggesting variability in the model's performance on the localization aspect over time. Notably, the mask loss, representing the model's segmentation capability, shows a steady and pronounced decline, signaling an enhanced proficiency in demarcating the affected tissue areas.

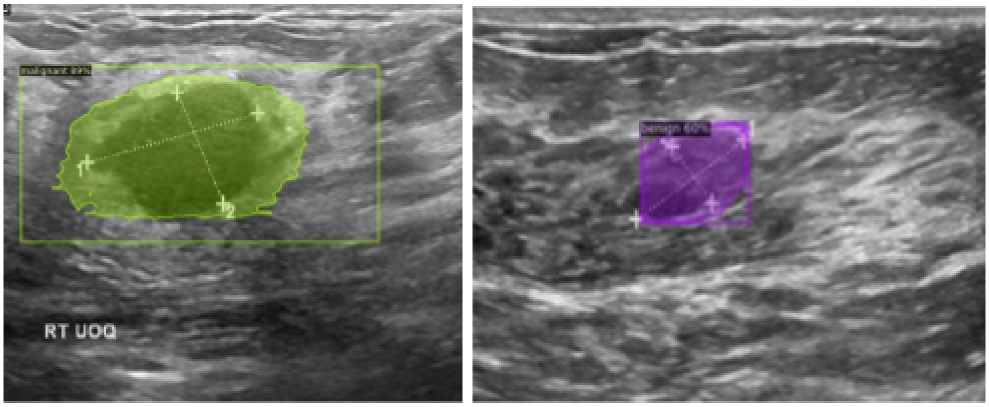
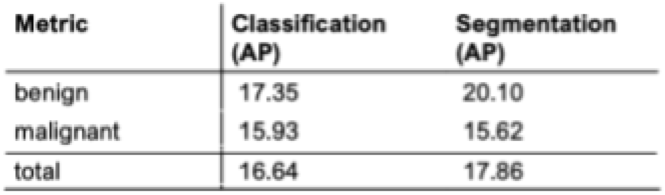


Figure 4: Discriminative Detection and Classification of Benign and Malignant in Sample Test Data Using Mask R-CNN

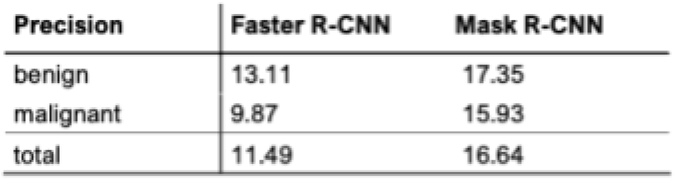
The performance metrics (Table 2) reveal a slightly better performance in correctly identifying benign instances compared to malignant one, also as Figure 4 suggests that the model is more adept at delineating the boundaries of benign features within the images than it is for malignant features. These results indicate that while the model is relatively effective at segmenting and classifying benign instances, there is room for improvement, especially in the accurate detection and segmentation of malignant cases.

In the context of classifying different lesion types on breast cancers, we examined the performance of two convolutional neural network architectures, Faster R-CNN and Mask R-CNN. Table 3 illustrates the Average Precision (AP) on Classification for both models. The



Metric	Classification (AP)	Segmentation (AP)
benign	17.35	20.10
malignant	15.93	15.62
total	16.64	17.86

Table 2: the Average Precision (AP) metrics for Mask R-CNN model



Precision	Faster R-CNN	Mask R-CNN
benign	13.11	17.35
malignant	9.87	15.93
total	11.49	16.64

Table 3: the Average Precision (AP) on Classification in Faster R-CNN and Mask R-CNN for Lesion Detection.

Mask R-CNN outperformed the Faster R-CNN across all categories, achieving an AP of 17.35 for benign lesions and 15.93 for malignant lesions, compared to 13.11 and 9.87, respectively, achieved by the Faster R-CNN. This enhancement reveals the effectiveness of masks, which may provide richer information for accurate classification in medical ultrasound images.

5.4 U-Net

We utilized the original architecture of the U-Net model consisting of 3 encoders each followed by max pooling, 2 decoders concatenated with the corresponding cropped feature map, and the final decoder. For each encoder and decoder, a series of two 3*3 convolutions, ReLU, and batch normalization are applied.

Since the provided ground truth data are binary masks and we preprocessed the data into gray-scale, we changed the activation function of the final layer into sigmoid and the loss function into Binary Cross Entropy (BCE) loss. Dropout 0.3 was added to each encoding and decoding block to prevent overfitting. During the training, the learning rate was adjusted from 0.005 to 0.01 to enhance the performance of the model faster. Figure 5 shows the performances of 2 models with different learning rates on validation data: ‘trial’ with 0.005, ‘final’ with 0.01.

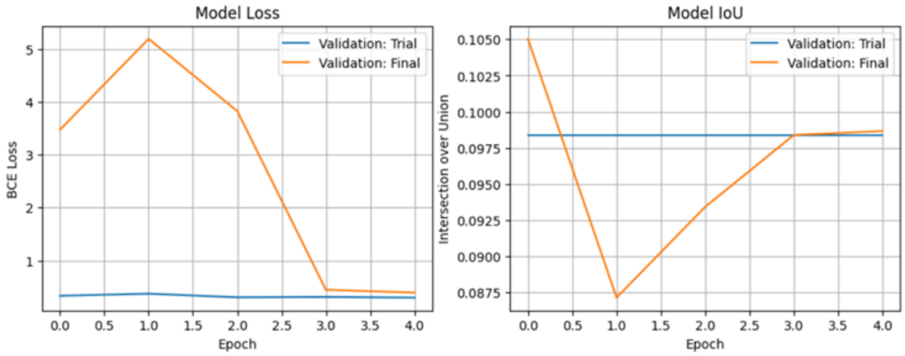


Figure 5. Binary Cross Entropy (BCE) loss and Intersection over Union (IoU) on training data

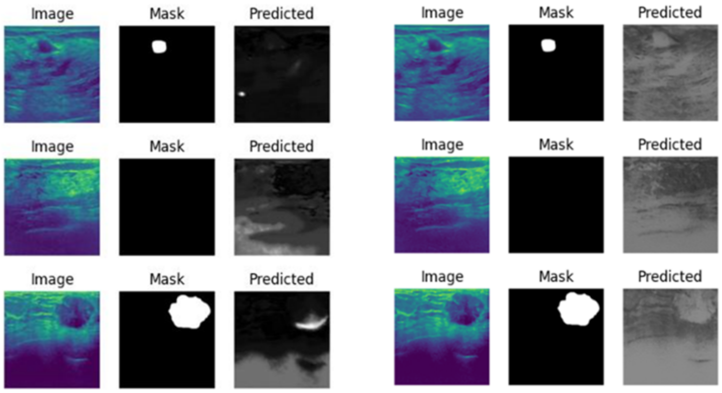


Figure 6. Segmentation predicted with different parameters on sample test data using U-Net
(column 1-3: after epoch 1 of trial model; column 4-6: after epoch 5 of final model)

For the efficiency of training, we set a stopping condition that if the average IoU on validation data doesn’t exceed the best IoU for 5 times, the model stops training. According to the stopping condition, the training stopped after epoch 5. On the completion of epoch 5, BCE loss and IoU on validation data were 0.3912 and 0.0987. On the test data, the average BCE loss was 0.6886, the average IoU 0.0925, and the precision 0.1181.

5.5 MedSAM

MedSAM is a pre-trained transformer model, so we fine-tuned the model with our dataset to improve the performance. We fine-tuned the model with batch size 1 and AdamW as an optimizer with 0.001 as a learning rate and 0.01 as weight decay for 5 epochs and Dice+ Cross entropy as a loss function. Dice+ Cross entropy was used as a loss function to cover both the spatial overlap which emphasizes generating accurate boundaries, and class imbalance to combat the nature of segmentation. The batch size of 1 is chosen because when the batch size was larger than 1, we encountered CUDA out-of-memory error when using Google Colab Pro's V100 and High-RAM. This exhibits the limitations of MedSAM and other vision transformer models that they are computationally expensive. Training the model to fine-tune based on the data took 285 minutes for 5 epochs, approximately 1 hour per epoch. The ramifications of such limitation are that not only it takes a significant time to train the model, hence only 5 epochs, but also could lead to overfitting and noisy gradation. As shown in the figure 6, it is observed that the loss generally trends downward but suddenly spikes at epoch 4.

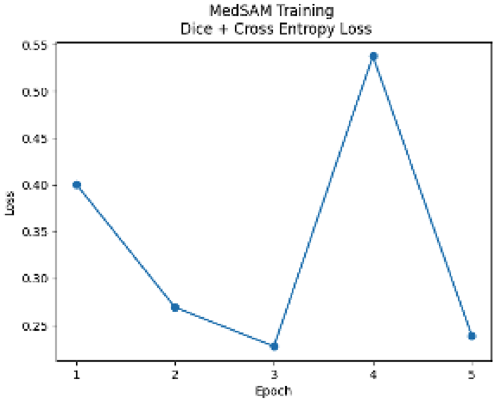


Figure 7: Dice + Cross Entropy loss of MedSAM training per epoch

After training to fine-tune the model, we compared the performance of the pre-trained MedSAM model with our fine-tuned MedSAM model using validation data to see how each model compares. Mean IoU of the pre-trained model on validation was 0.8761 and the fine-tuned mean was 0.6578. We also compared the result qualitatively by examining the output segmentation. The segmentation of the fine-tuned model covers more grounds that the pre-trained model missed, but it also overestimates the area of legions. Moreover, the shape of the pre-trained model looks more natural compared to the result of the fine-tuned model, which has rigid edges. The mean IoU for the test data is 0.6593, which is comparable to the result from validation. However, considering that the only a relatively small fine-tuning was done with the batch size of 1 which would have contributed to overfitting, the result might be better if there were more generalized fine-tuning would have been done if the resources permit.

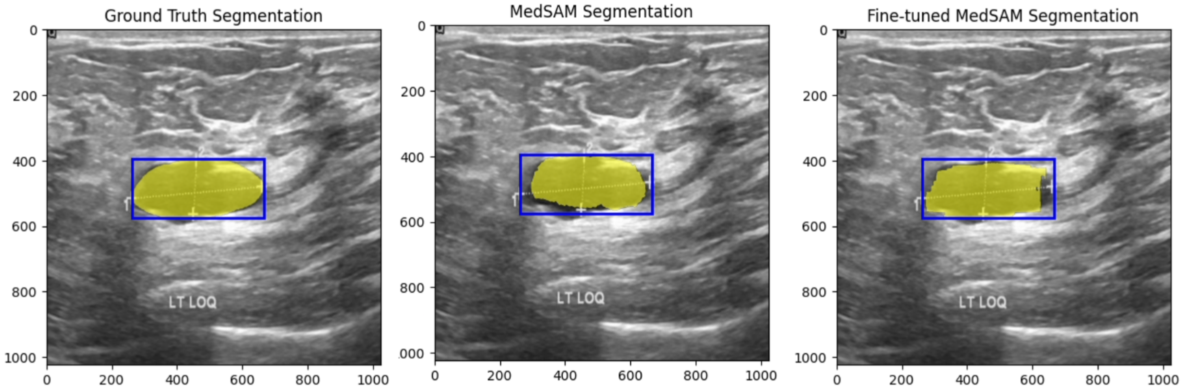


Figure 8 Comparison of ground truth, pre-trained model and fine-tuned model segmentation

6. Discussion and Conclusion

In this project, we implemented various image segmentation models to segment lesions of breast ultrasound to track the development of lumps. Table 4 shows the performance of each model in terms of average precision (AP) and intersection over union (IoU).

	Mask R-CNN	U-Net	Med-SAM
Average Precision	17.86	0.12	65.93
Intersection over Union	16.64	0.09	87.61 (pre-trained) 65.78 (fine-tuned)

Table 4. Performance of Segmentation Models

In table 4, Med-SAM showed the highest evaluation score, whereas U-Net showed the lowest. Since Med-SAM is a pre-trained model based on millions of biomedical images, it also showed better performance on the breast ultrasound images. For U-Net on the other hand, because of its natural contracting structure, the small lesions on the image can vanish during the convolutional process. For that reason, some significant insights on the application of U-Net was derived: detection models preceding U-Net can significantly enhance the performance; stopping criteria should be carefully considered when the region of interest is relatively small.

During the implementation, we noticed major challenges. One challenge was difficulty in training. It took a long time to train the models, and the resources we have available were not enough to execute expensive computation. It took U-Net 2 hours per epoch, MedSAM 1 hour per epoch for training. That resulted in poorer performance on fine-tuned MedSAM that the pre-trained MedSAM model had better performance on validation data than the fine-tuned model (pre-trained mean IoU 87.61 vs fine-tuned mean IoU 65.78). We hope to be able to examine in the future whether the performance would be improved if we have access to better computation power. Other challenge that we faced was based on the nature of medical images, especially due to small dataset and images with noise and low contrast. We augmented the data to enhance the training process and promote generalization, and we were able to achieve adequate metrics. Based on our experiments, exactly what was the cause of the relatively lower performance is unknown. It would most likely due to both the data quality and quantity, and the computation resource.

In the field of medicine, where accuracy is extremely important and inaccurate data can pose the highest risk among the application of machine learning, adequate performance might not be enough. For a practical implementation, it might be better to be conservative and use a simpler model to merely detect lesions, so it can assist the medical professionals in deciding diagnosis rather than implementing complex models with risk.

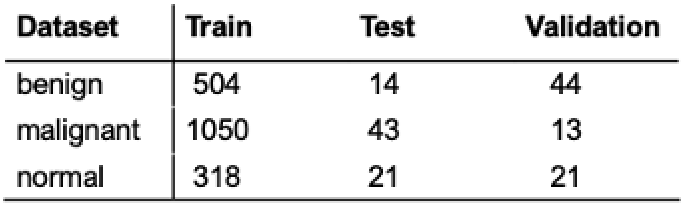
7. Individual Contributions

Sohyun Shin: I was in charge of implementing and evaluating the MedSAM model. I wrote the experiments and evaluation part for MedSAM and also the introduction and related works, and the part for discussion and conclusion.

Xinyu Zhang: I did data preprocessing, which encompasses data splitting and augmentation. Additionally, I employ Mask R-CNN and Faster R-CNN models for classification, object detection, and segmentation tasks on the ultrasound image dataset.

Yujeong Yang: I implemented and evaluated the U-Net model, wrote the part of experiments and evaluation on U-Net, and the discussion and conclusion part.

Appendix



Dataset	Train	Test	Validation
benign	504	14	44
malignant	1050	43	13
normal	318	21	21

Table 1: Class Distribution across Training, Testing, and Validation Sets.

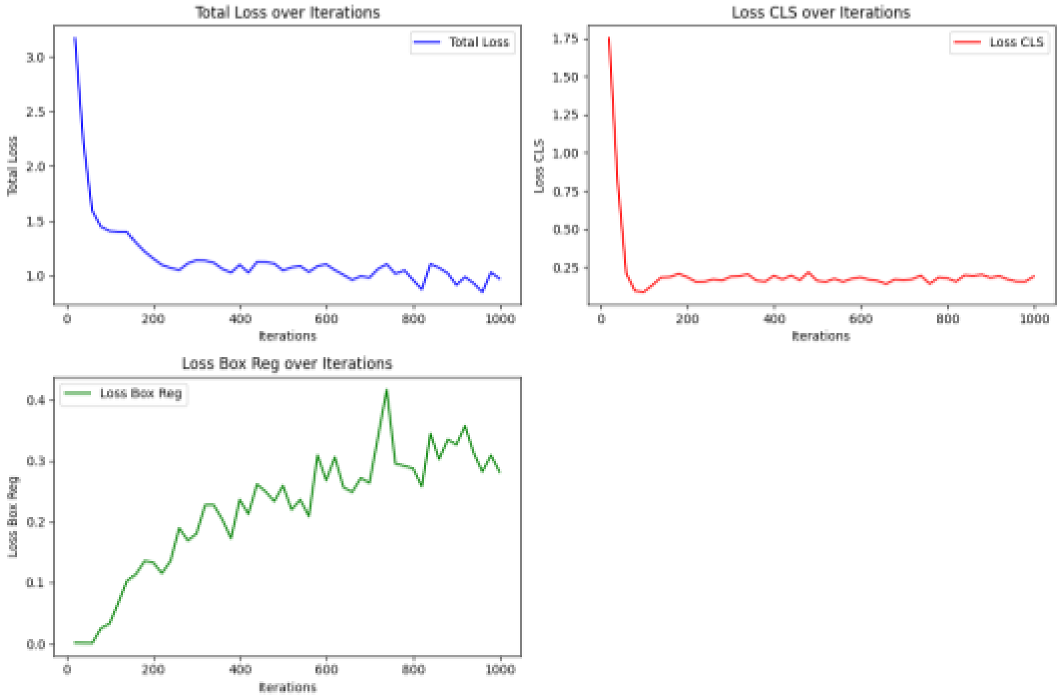


Figure 2: Training Loss Metrics Over iterations for Faster R-CNN model

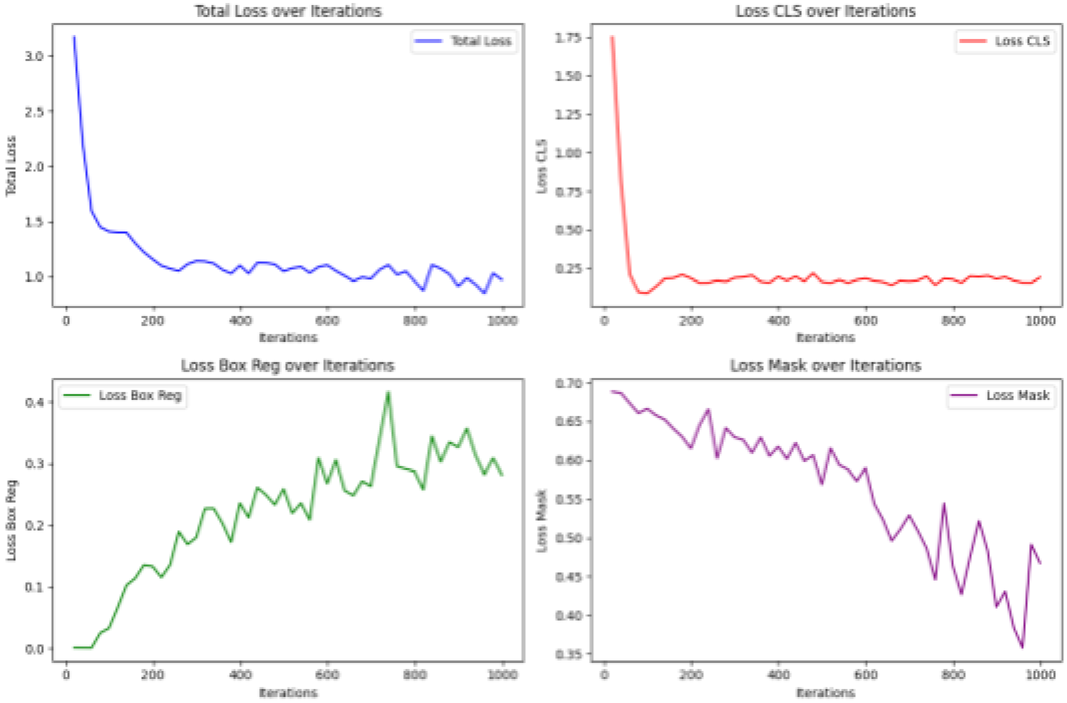


Figure 3: Training Loss Metrics Over iterations for Mask R-CNN model

Citation

[1] American Cancer Society, “Breast Cancer Statistics | How Common Is Breast Cancer?,” www.cancer.org, Jan. 12, 2023. https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html

[2] American Cancer Society, “ACS breast cancer screening guidelines,” www.cancer.org, Jan. 14, 2022. https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/american-cancer-society-recommendations-for-the-early-detection-of-breast-cancer.html

[3] M. F. Mridha et al., “A Comprehensive Survey on Deep-Learning-Based Breast Cancer Diagnosis,” Cancers, vol. 13, no. 23, p. 6116, Dec. 2021, doi: https://doi.org/10.3390/cancers13236116.

[4] M. Hu, Y. Li, and X. Yang, “BreastSAM: A Study of Segment Anything Model for Breast Tumor Detection in Ultrasound Images” arXiv.org, May 21, 2023. https://arxiv.org/abs/2305.12447 (accessed Oct. 23, 2023).

[5] Muhammad Azaz Farooq, Z. Gong, Y. Liu, M. Zubair, A. Manzoor, and G. Zhang, “Breast cancer detection from ultrasound images using attention U-nets model” Fourteenth International Conference on Digital Image Processing (ICDIP 2022), Oct. 2022, doi: https://doi.org/10.1117/12.2643599.

[6] M. Ahmadi et al., “Comparative Analysis of Segment Anything Model and U-Net for Breast Tumor Detection in Ultrasound and Mammography Images” arXiv.org, Jun. 21, 2023. https://arxiv.org/abs/2306.12510 (accessed Sep. 07, 2023).

[7] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, “Dataset of breast ultrasound images,” Data in Brief, vol. 28, p. 104863, Feb. 2020, doi: https://doi.org/10.1016/j.dib.2019.104863.

[8] J. Ma and B. Wang, “Segment Anything in Medical Images,” arXiv.org, Apr. 24, 2023. https://arxiv.org/abs/2304.12306

[9] S. Gokhale, “Ultrasound characterization of breast masses,” Indian Journal of Radiology and Imaging, vol. 19, no. 3, p. 242, 2009, doi: https://doi.org/10.4103/0971-3026.54878.

[10] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28