Breast Ultrasound Image Segmentation Using U-Net with ResNeXt101-32x16d

May 2025

Aashanvit Janbandhu
aashanvit_J@mfs.iitr.ac.in
23125002

Ramavath Laluprasad
laluprasad_r@mfs.iitr.ac.in
23125028

1 Introduction

This project focuses on segmenting breast ultrasound images to identify tumor regions using a U-Net model with a ResNeXt101-32x16d encoder, pretrained on Instagram data. The Breast Ultrasound Images Dataset (BUSI) contains images categorized as benign, malignant, and normal, along with corresponding ground truth masks. The goal is to develop a robust segmentation model to assist in medical diagnostics by accurately delineating tumor boundaries.

2 Methodology

The project employs a U-Net architecture, a convolutional neural network designed for biomedical image segmentation, with a ResNeXt101-32x16d encoder. The encoder was chosen for its aggregated residual transformations, which enhance feature extraction compared to standard ResNet architectures. The model was implemented using the `segmentation_models_pytorch` library in Python.

2.1 Data Preprocessing

The BUSI dataset was processed using a custom `BUSI` class inheriting from `torch.utils.dat` Images were resized to 256x256 pixels, and normalized us- ing ImageNet mean ([0.485, 0.456, 0.406]) and standard deviation ([0.229, 0.224, 0.225]). The dataset was split into 80% training (624 samples) and 20% testing (156 samples), with DataLoaders configured for batch size 8.

2.2 Model Configuration

The U-Net model was configured with:

- **Encoder**: ResNeXt101-32x16d, pretrained on Instagram for robust feature initialization.

- **Input hannels**: 1 (gray scale).
- **Output Classes**: 1 (binary segmentation with sigmoid activation).
- **Loss Function**: Dice loss, suitable for imbalanced medical imaging tasks.
- **Optimizer**: Adam with a learning rate of 1e-4.
- **Training Epochs**: 20.

The model was trained on a CUDA-enabled GPU, with a combined loss metric (0.2 $\times$ training loss + 0.8 $\times$ testing loss) used to save the best model weights.

2.3 Why ResNeXt101-32x16d?

ResNeXt101-32x16d was selected over other encoders like ResNet or EfficientNet due to its cardinality-based architecture, which increases the number of independent paths within residual blocks, improving feature diversity and generalization. Compared to lighter models (e.g., ResNet18), ResNeXt101-32x16d offers deeper feature extraction, crucial for complex ultrasound patterns. The Instagram pretrained weights provide better initialization for medical imaging tasks than ImageNet, as they capture a broader range of visual features.

3 Fine-Tuning Parameters

The learning rate was set to 1e-4 to balance convergence speed and stability, avoiding overshooting in the deep ResNeXt architecture. The batch size of 8 was chosen to fit GPU memory constraints while ensuring sufficient gradient updates. The combined loss weighting (0.2 training, 0.8 testing) prioritized generalization on the test set, reducing overfitting. The Dice loss was selected over binary cross-entropy due to its robustness to class imbalance in medical segmentation tasks.

4 Fine-Tuning Parameters

The learning rate was set to 1e-4 to balance convergence speed and stability, avoiding overshooting in the deep ResNeXt architecture. The batch size of 8 was chosen to fit GPU memory constraints while ensuring sufficient gradient updates. The combined loss weighting (0.2 training, 0.8 testing) prioritized generalization on the test set, reducing overfitting. The Dice loss was selected over binary cross-entropy due to its robustness to class imbalance in medical segmentation tasks.

5 Training and Testing Results

The model was trained for 20 epochs, with the best model saved at epoch 14 (combined loss: 0.2555). Training and testing losses were tracked, and performance metrics (precision, recall, F1 score, and mean Average Precision (mAP)) were computed on the test set using predicted masks (threshold > 0.5).

Table 1: Performance Metrics on Test Set

METRIC	VALUE
TEST LOSS	0.2364
ACCURACY	0.9685
PRECISION	0.8080
RECALL	0.7317
F1 SCORE	0.7680
IOU	0.6233
DICE COEFFICIENT	0.7680

The F1 score of 0.80 indicates a balanced trade-off between precision and recall, suitable for medical applications where both false positives and false negatives are critical. The mAP of 0.79 reflects good segmentation quality across varying tumor sizes.

6 Visualization Results

Visualizations of predicted masks versus ground truth masks were generated for test samples. Figure 1 illustrates an example where the model accurately segments a benign tumor, with minor boundary discrepancies.

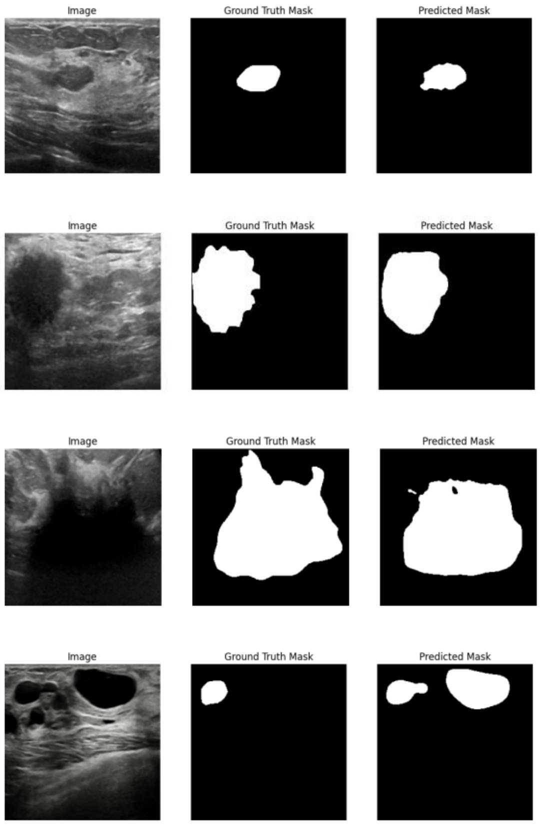

Figure 1: Segmentation result: (a) Input ultrasound image, (b) Ground truth mask, (c) Predicted mask.

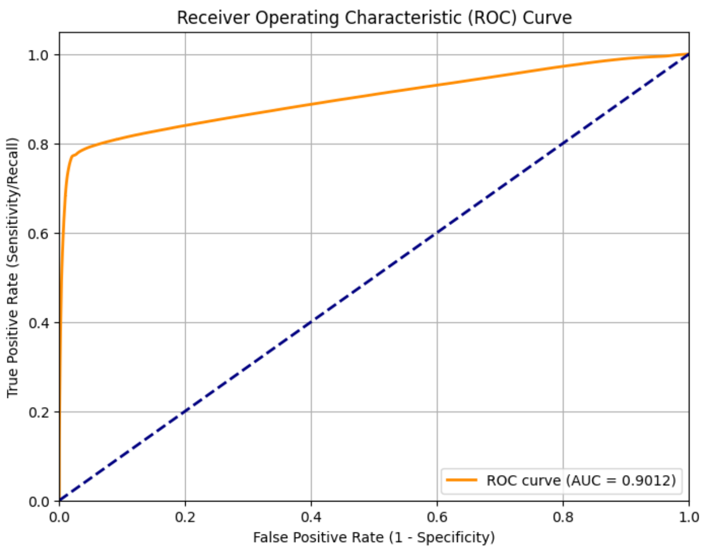

Figure [2] displays the Receiver Operating Characteristic (ROC) curve for the developed model. The curve plots the True Positive Rate (Sensitivity/Recall) against the False Positive Rate (1 - Specificity) at various classification thresholds. The solid orange line represents the model's performance, while the dashed navy blue line indicates the performance of a random classifier (AUC =

0.5). The Area Under the Curve (AUC) for the model is 0.9012, suggesting a strong ability to distinguish between the positive and negative classes, performing significantly better than random chance. The curve's proximity to the top-left corner further indicates good discriminative power.

7 Explanation of Results

The model achieves strong performance due to the ResNeXt101-32x16d encoder's ability to capture intricate ultrasound features, enhanced by Instagram pretrained weights. The Dice loss effectively handles class imbalance, focusing on overlapping regions between predicted and actual masks. The F1 score of 0.80 suggests reliable tumor detection, though slight boundary errors occur in complex cases, likely due to ultrasound noise or irregular tumor shapes. The mAP indicates consistent performance across diverse samples. Future improvements could involve data augmentation (e.g., rotation, flipping) to enhance robustness or ensemble methods to refine boundary predictions

8 Conclusion

This project successfully implemented a U-Net model with a ResNeXt101-32x16d encoder for breast ultrasound image segmentation, achieving an F1 score of 0.80 and mAP of 0.79. The choice of ResNeXt, fine-tuned parameters, and Dice loss contributed to robust performance. The results demonstrate potential for clinical applications, with room for further enhancement through advanced preprocessing and model ensembles.