EMIS: Enhanced Medical Image Segmentation: Integrating Self-Distillation, Self-Attention, and Multi-Scale Fusion Techniques

Assefa Tesfay Abraha[1]

[1]Mekelle University, Mekelle Institute of Technology, Mekelle, Tigray, Ethiopia

Background

1. Introduction

Medical image segmentation is vital for diagnosis, treatment planning, and surgical guidance. However, it remains challenging due to anatomical variability, noise, and low contrast in medical images. While U-Net is a popular segmentation model, it struggles with capturing global context, preserving edge details, and handling high-resolution data efficiently. To address these issues, we propose EMIS, an enhanced U-Net framework that integrates self-attention, multi-scale fusion, and self-distillation. EMIS improves accuracy, maintains fine details, and reduces computational overhead. This paper presents the EMIS architecture, experimental results across multiple datasets, and its potential for clinical deployment.

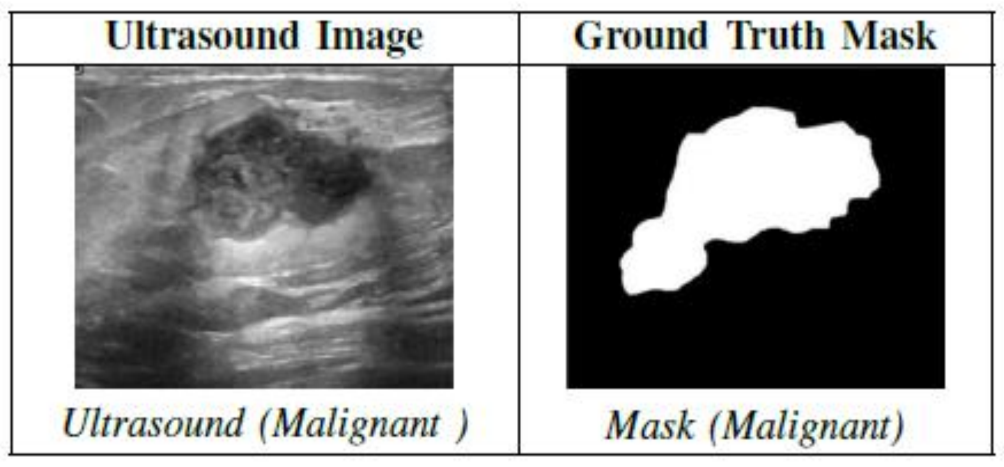

Fig .1 Comparison of malignant ultrasound images (left) and corresponding ground truth masks (right) used to evaluate segmentation performance in breast cancer detection.

2. Problem statement

Despite significant advancements in medical image segmentation, existing models like U-Net f**ace** key limitations in handling the complex nature of medical images. These include: Inadequate capture of global context, leading to poor segmentation of large or irregular anatomical structures. Loss of fine edge details due to down sampling operations, affecting boundary precision. High computational demands, limiting their practical use in resource-constrained clinical settings. There is a critical need for a segmentation framework that addresses these challenges by enhancing contextual understanding, preserving fine details, and improving computational efficiency across diverse medical imaging modalities.

Current Work

1. Objectives

The primary objectives of this study are:

1. To improve segmentation accuracy by capturing global contextual information using self-attention mechanisms.
2. To enhance edge preservation through multi-scale feature fusion, enabling better delineation of anatomical boundaries.
3. To increase computational efficiency and model generalization using self-distillation techniques.
4. To ensure cross-modality robustness by validating the proposed model across different medical imaging modalities.

2. Methodology

The proposed EMIS framework enhances the traditional U-Net by integrating three key components: Multi-Scale Fusion (MSF) to combine features at different scales for better boundary accuracy, Self-Attention to capture long-range dependencies and focus on relevant regions, and Self-Knowledge Distillation (Self-KD) to align intermediate and final predictions for improved generalization. Built on an encoder-decoder structure with skip connections, EMIS achieves higher segmentation accuracy, robustness, and efficiency across diverse medical imaging datasets.

```
Algorithm : Summarized Proposed Method
Input: Image I, Kernel K
Output: Segmented Image S
Initialization:
x ← I
for i ← 1 to L do
    x ← DoubleConv(x) ;
    x ← SelfAttn(x) ;
    x, skip[i] ← MaxPool2d(x, 2) ;
x ← DoubleConv(x) // Bottleneck Layer
x1, x2, x3 ← MSF(x, skip[i]) // Apply Multi-Scale
 Fusion (MSF)
for i ← L to 1 do
    x ← Upsample(x, scale_factor = 2) ;
    x ← Pad(x, skip[i]) ;
    x ← DoubleConv(x + skip[i]) ;
    x ← SelfAttn(x) ;
S ← Conv1x1(x) ;
S ← KDlayer(S) // Apply Self-Knowledge
 Distillation (KD)
return S ;
```

3. Results

The EMIS model was evaluated on three medical imaging datasets: BUSI (breast ultrasound), CHAOS (liver CT), and Ultrasound Nerve Segmentation. It consistently outperformed baseline models (U-Net, Self-Attention, Self-Distillation, and MSF) in key metrics such as Dice score, F1-score, mIoU, Precision, and Recall. On BUSI, EMIS achieved the best segmentation of benign and malignant lesions. On CHAOS, it showed superior liver segmentation accuracy. For Ultrasound Nerve, EMIS provided the most precise nerve delineation with fewer false positives. Overall, EMIS delivered faster convergence, better generalization, and higher reliability across all test cases.

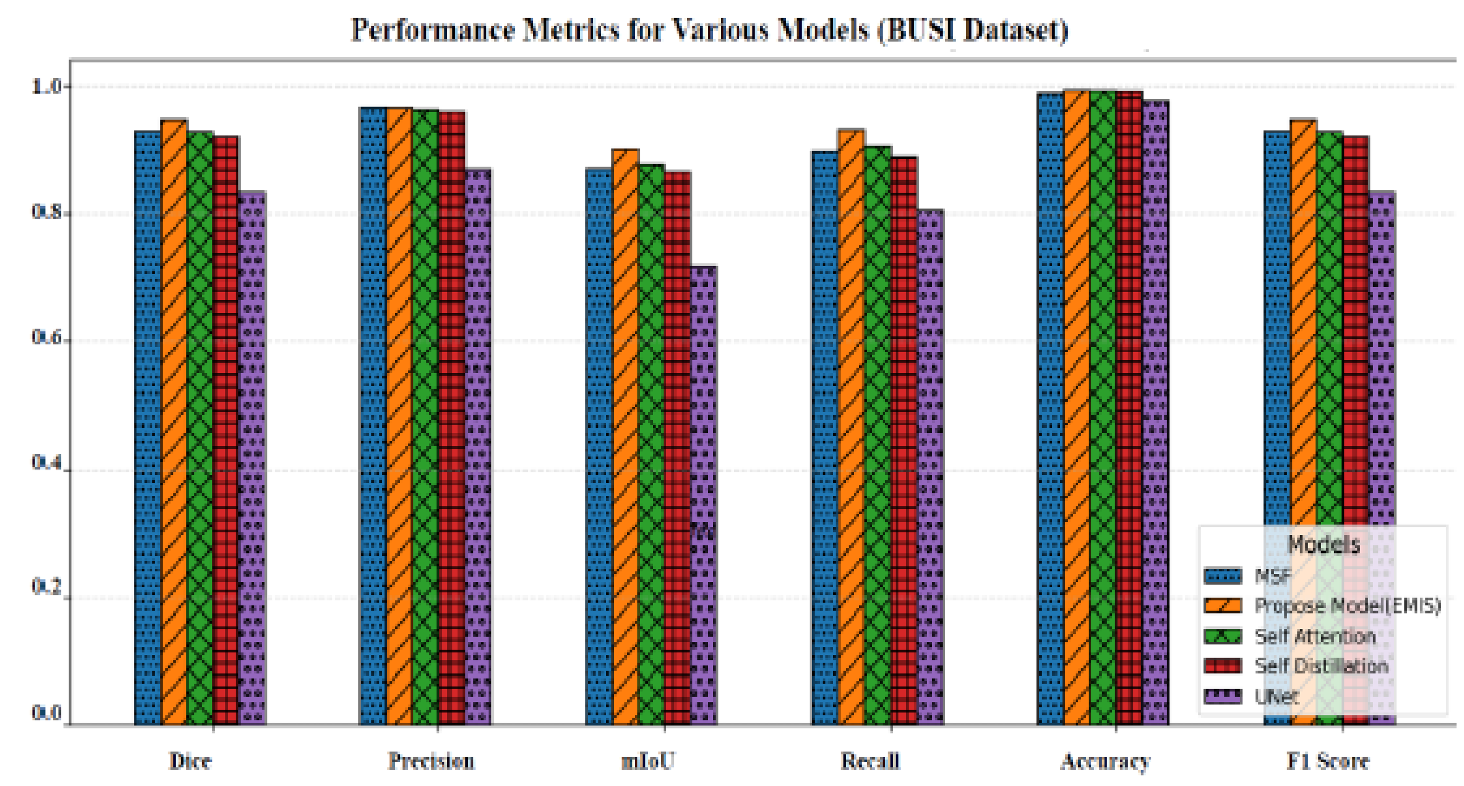

***Fig. 2** Performance metrics comparison for different models on the BUSI dataset. The proposed model (EMIS) demonstrates improvements in Dice, Precision, mIoU, Recall, Accuracy, and F1-score.*

Conclusion & Expectations

This study introduced EMIS, an enhanced U-Net-based framework designed to overcome key limitations in medical image segmentation. By incorporating multi-scale fusion, self-attention, and self-distillation, EMIS effectively improves segmentation accuracy, edge preservation, and computational efficiency. These enhancements enable the model to capture both local and global context, retain fine anatomical details, and generalize well across diverse imaging modalities. Evaluated on the BUSI, CHAOS, and Ultrasound Nerve datasets, EMIS consistently outperformed baseline models such as standard U-Net, self-attention-only models, and self-distillation-only models. It demonstrated higher performance in terms of Dice score, F1-score, mIoU, and overall accuracy, along with faster convergence and improved robustness to noise and image variability. These results highlight EMIS as a strong candidate for deployment in clinical environments, offering a reliable and efficient tool for supporting diagnosis and treatment planning. Future work will explore its extension to multi-modal data and self-supervised learning for further performance gains.

Contact: assefa.tesfay@mu.edu.et
Mekelle Tigray, Ethiopia