Real-Time Segmentation of Ultrasonic Images on Mobile Devices

Rikhat Akizhanov, Nuren Zhaksylyk, Aidar Myrzakhan
Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE
{rikhat.akizhanov, nuren.zhaksylyk, aidar.myrzakhan}@mbzuai.ac.ae
Group ID = G-20

Abstract

This project aims to advance medical imaging by creating a lightweight, efficient AI model for mobile devices. The goal is to integrate ultrasonic technology, mobile devices, and neural networks to offer revolutionary, accessible, and user-friendly diagnostic tools. The focus is on designing neural network models for real-time instance segmentation on ultrasonic images. We reviewed several lightweight models that demonstrate high performance and low latency, making them suitable for real-time inference in mobile-based health applications. Further we evaluated and tested their speed performance on demo mobile application taking video input from real ultrasonic device.
Code is available at GitHub repository.

1 Introduction

The project addresses the critical need for swift and accurate medical diagnostics accessible through portable devices. The primary objective is to develop a robust framework capable of segmenting ultrasonic images in real-time on mobile platforms, overcoming constraints like limited computing power and storage.

The motivation for this project stems from the growing demand for mobile healthcare solutions that can deliver immediate, on-site diagnostic insights, particularly in remote or underserved areas. Ultrasonic imaging is a non-invasive, cost-effective diagnostic tool widely used in various medical fields. However, its efficacy heavily relies on the quality of image interpretation, which can be challenging and time-consuming. By leveraging advanced machine learning models, this project aims to facilitate faster, more accurate interpretations directly on mobile devices, thus democratizing access to essential healthcare services.

To achieve this, we explored various deep learning models known for their efficiency and accuracy in image segmentation tasks. The selection included SegResNet [12], MobileNetV3[6], ResUNet[4], MobileViTV2[11], and Segformer[15], all benchmarked on their ability to segment ultrasonic images using the DICE score metric. We particularly focused on optimizing the models for real-time performance, assessing their speed in frames per second (FPS) to ensure their viability for mobile deployment. The datasets used for training and evaluating these models were the Breast Ultrasound Images (BUSI)[3], Open Kidney Ultrasound[13] and CT to Ultrasound (CT2US)[14], chosen for their relevance and diversity in ultrasonic imaging scenarios.

Contributions

Our contributions can be summarized as follows.

- Train different lightweight segmentation models on ultrasonic datasets.

- Evaluate and compare the models with baselines and assess their effectiveness on mobile devices.
- Develop simple yet demonstrative mobile application for segmentation of medical images on Android devices using PyTorch and Java.
- Demonstrate effectiveness and feasibility of real-time segmentation on ultrasonic images.

2 Related Work

Recent advancements in AI have significantly impacted medical imaging, enabling more precise and faster diagnostics. The trend towards mobile healthcare necessitates efficient AI models. SegResNet, introduced by Myronenko [12], integrates the strengths of deep residual networks and segmentation-based networks for medical imaging analysis. Its architecture, tailored for 3D imaging data, has shown significant effectiveness in segmenting intricate structures in medical images, particularly in brain MRI analysis. The model's depth and residual connections contribute to its robust performance in capturing fine-grained details, essential for accurate medical diagnostics.

Developed by Howard et al.[6], MobileNetV3 stands out for its efficiency and effectiveness in mobile applications. This model leverages an architecture optimized through hardware-aware neural architecture search (NAS), making it highly suitable for deployment on mobile devices with limited computational resources. Its use in medical imaging signifies a step forward in facilitating accessible, real-time diagnostics on portable devices.

Diakogiannis et al.[4] proposed ResUNet, a model combining the strengths of Residual Networks and U-Net. It is particularly designed for semantic segmentation tasks with a focus on enhancing feature extraction while maintaining computational efficiency. In medical imaging, ResUNet has been effective in segmenting complex anatomical structures, offering a balance between depth and performance.

Mehta et al.'s [11] MobileViTV2 represents a novel approach by integrating Vision Transformers into a mobile-friendly architecture. This model addresses the need for high accuracy in image analysis while being computationally efficient for mobile deployment. Its application in medical imaging shows promising results, especially in scenarios where detailed image analysis is required on handheld devices.

Segformer, introduced by Xie et al. [15], is a unique segmentation model that combines a hierarchical transformer with a simple decoder to achieve high efficiency and accuracy. Its design is particularly suitable for processing large-scale medical images, providing detailed segmentation results crucial for accurate medical diagnostics.

In evaluating the state-of-the-art, we consider the comprehensive analysis by [8] and [14], which propose novel methods in medical imaging. These studies set a benchmarks for model performance, against which our project's outcomes are measured. Our approach differs from these models by focusing on real-time processing capabilities on mobile platforms, addressing a gap in immediate diagnostic requirements.

Our project intersects these key areas, contributing to the field by developing efficient AI models suitable for mobile devices. By utilizing and extending the capabilities demonstrated in existing datasets like BUSI[3] and CT2US[14], we aim to push the boundaries of what's achievable in mobile medical diagnostics. Our unique contribution lies in the adaptation and optimization of lightweight models, specifically tailored for real-time, on-device medical image analysis.

3 Problem Statement

The central issue addressed in this project is the challenge of implementing real-time, accurate ultrasonic image segmentation on mobile devices, a task traditionally hindered by the limited computational resources of these platforms. This innovation is crucial for enhancing medical diagnostics in remote and underserved areas, where access to expert analysis and advanced medical equipment is often scarce. The project aims to leverage and optimize state-of-the-art deep learning models like SegResNet, MobileNetV3, ResUNet, MobileViTV2, and Segformer for this purpose, focusing on their efficiency in segmenting ultrasonic images. The models are evaluated based on the DICE score metric and their processing speed, ensuring their practicality for mobile healthcare applications. The project utilizes diverse and relevant datasets such as BUSI, Open Kidney Ultrasound, and CT2US, reflecting the variety of scenarios in ultrasonic medical imaging.

4 Methodology

4.1 Models

The chosen models for this project, SegResNet, MobileNetV3, ResUNet, MobileViTV2, and Segformer, were selected based on their unique attributes suitable for ultrasonic image segmentation on mobile platforms:

1. SegResNet[12]: Known for its effectiveness in 3D imaging, SegResNet combines deep residual learning with segmentation capabilities, making it ideal for detailed anatomical analysis in ultrasonic images.
2. MobileNetV3[6]: Optimized for mobile devices, MobileNetV3 offers a balance between efficiency and accuracy, crucial for real-time applications in mobile healthcare.
3. ResUNet[4]: This model merges the robust feature extraction of Residual Networks with the segmentation prowess of U-Net, suitable for precise segmentation tasks.
4. MobileViTV2[11]: Integrating Vision Transformers into a mobile-friendly architecture, MobileViTV2 is chosen for its ability to handle detailed image analysis efficiently on mobile devices.
5. Segformer[15]: With its hierarchical transformer structure, Segformer is adept at processing large-scale images, providing high-quality segmentation which is essential for accurate medical diagnostics.

These models collectively offer a comprehensive approach to tackling the challenges of mobile-based ultrasonic image segmentation, balancing speed, accuracy, and computational efficiency.

4.2 Data Preparation

In the preprocessing phase, several steps were implemented to prepare the datasets for effective model training. To enhance model robustness, the following random transformations were applied:

- Horizontal and vertical flipping with probabilities of 0.5
- Rotation between -30 and 30 degrees.
- Affine transformations including rotation, translation, scaling, and shearing.
- Resized cropping to a uniform size with anti-aliasing.

These steps were crucial for normalizing the data and enhancing the models' ability to generalize from the training data.

4.3 Optimization

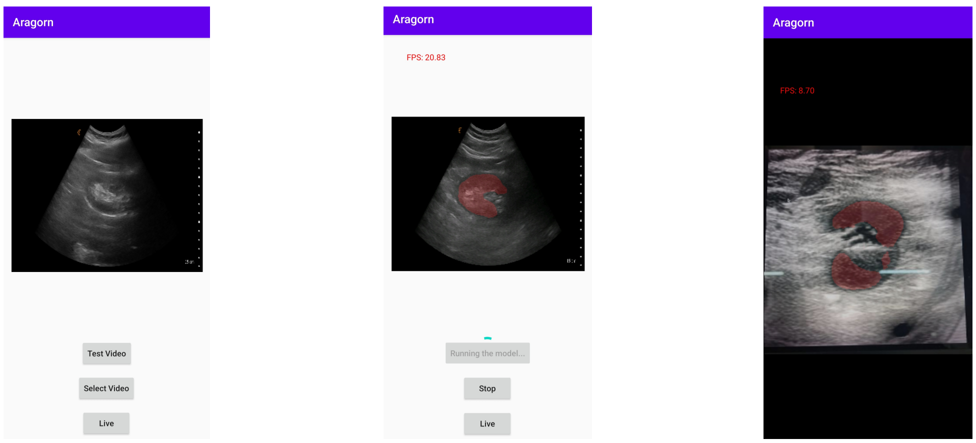
AdamW Optimizer[10]: a modification of the Adam optimizer with decoupled weight decay regularization, was employed. This optimizer is known for effectively handling sparse gradients and providing stable training, crucial in medical imaging contexts.

DiceCE Loss[5]: function, combining Dice loss and Cross-Entropy loss, was used to address class imbalance issues in medical image segmentation. Dice loss ensures a high overlap with the ground truth, while Cross-Entropy loss improves pixel-wise classification accuracy.

These strategies, along with tailored hyperparameters, were key in enhancing model efficiency and viability for real-time applications on mobile devices.

4.4 Mobile Application

Our application represents a leap forward in the field of medical imaging, particularly in the analysis of ultrasonic images. The app is intuitive yet powerful, and equipped with features to make segmentation on video and directly from stream taken from ultrasonic device. Also, the speed of models can be seen on every frame running.
Test Video: This feature enables the analysis of pre-recorded ultrasonic videos. The application processes each frame sequentially, providing detailed segmentation of the ultrasonic images.
Select Video: This functionality allows for the importation and subsequent analysis of ultrasonic videos from the user's device.
Live: Designed for real-time analysis, this feature processes live ultrasonic imaging streams from the device.



(a) Main Page (b) Test Video Demonstration (c) Live Demonstration

5 Experimental Setup

5.1 Datasets

The project utilized three diverse and significant datasets, each contributing uniquely to the field of ultrasonic image analysis.

The first dataset, BUSI (Breast Ultrasound Images)[3], is pivotal for breast cancer detection. It comprises 780 ultrasound images with dimensions of 500x500 pixels, collected from 600 female patients aged between 25 and 75 years in 2018. The dataset categorizes images into normal, benign, and malignant classes, thus providing a comprehensive basis for training models in breast cancer classification, detection, and segmentation.

The second dataset, CT2US[14], addresses the challenge of kidney segmentation in ultrasound imaging. This dataset employs a novel approach of cross-modal transfer learning, where annotated data from CT scans are leveraged to enhance ultrasound image analysis. By utilizing CycleGAN for synthesizing ultrasound images from CT data, the dataset effectively mitigates the domain discrepancy between the two imaging modalities. This approach not only improves segmentation accuracy but also enhances the generalization capabilities of the models with limited ultrasound training data.

Finally, the Open Kidney Dataset[13] stands as the first publicly available dataset for kidney B-mode ultrasound data. It includes data retrospectively collected over a five-year period from more than 500 patients, encompassing both native and transplanted kidneys. The dataset features fine-grained manual annotations from two expert sonographers and labels for various views, providing an invaluable resource for developing novel image analysis techniques in kidney disease detection and prognosis. With its diverse range of images and expert annotations, the Open Kidney Dataset offers a high-quality basis for advanced ultrasound image analysis.

5.2 Evaulation Metric(s)

For evaluating the models, we focused on two key metrics:

DICE Score: This metric measures the similarity between the predicted segmentation and the ground truth, making it highly relevant for assessing the accuracy of medical image segmentation models. It is defined as:

$$\text{DICE Score} = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$

where X and Y represent the predicted and ground truth segmentation, respectively.

Frames Per Second (FPS): This metric evaluates the model's speed, crucial for real-time applications on mobile devices. A higher FPS indicates faster processing, essential for on-site, immediate diagnostic insights.

These metrics were carefully chosen to ensure that the models are not only accurate but also practical for use in mobile healthcare scenarios.

5.3 Training

All training was performed on single Nvidia Quadro RTX 6000 GPU. To ensure convergence of the models we increased number of epochs to be 80 for all experiments. We started with learning rate equal to 0.001 and then used Cosine Annealing[9] scheduler to gradually learning rate and get smoother training process. Small batch size can give regularization effect[7] to the model which is especially beneficial in medical imaging where not much data is available. In this way batch size of 16 was chosen and used throughout the experiment. Higher image size can increase prediction score as model will get more information and on the other hand can increase inference time, thus medium value of 256x256 was selected to accurate enough results and have fast models.

5.4 Mobile Deployment

For our project we are about to use Clarius C3 HD_3 [1] multipurpose ultrasonic device. It operates in the frequency of 2-6 MHz and has max depth of 40 cm. It is wireless and has API to access the images in real-time on the screen of the mobile device.

In our research, we utilized Android Studio and Java for the development of our mobile application. Also, we used PyTorch Mobile packages [2] to enable on-device machine learning model inference. In this task, we tested the application on 2 Android devices: Samsung S20 with Exynos 990 and 8GB RAM; Xiaomi Mi Note 10 Pro with Snapdragon 730G and 8GB RAM.

6 Results and Discussion

Our experimental evaluations on the BUSI, CT2US, and Open Kidney datasets have yielded insightful results regarding the efficacy of various deep learning models in ultrasonic image segmentation. These results are captured quantitatively by the Dice score metric and qualitatively through visual inspection of the segmented images.

Overall selected models achieve reasonably high results, demonstrating that they can find and segment object on ultrasonic images. Comparing to related works on same datasets, we achieve same or better results on CT2US dataset as comparing to [14]. Nevertheless, real data can be noisy and harder to segment with high precision, results of [8] higher than our on BUSI, but it should be mentioned that they utilize much larger networks to solve this problem.

SegResNet demonstrated a remarkable balance between efficiency and accuracy, particularly excelling in the Open Kidney dataset with a Dice score of 0.889 and the lowest parameter count. MobileNetV3, with its superior performance on the CT2US and BUSI datasets, reaffirmed its robustness across different segmentation scenarios. ResUNet and MobileViTV2 offered competitive results, especially in the complex segmentation tasks evidenced in the Open Kidney dataset. Segformer, while slightly trailing in Dice scores, maintained commendable accuracy across all testing scenarios.

Taking into account results of all the models, we can conclude that SegResNet is the best choice for our application, based on the fact that it achieves top results or very close to top models, while having drastically less parameters and hence will have singificantly lower inference time.

The Table 1 summarizes the comparative analysis of these models across various metrics.

Table 1: Comparison of DICE Scores, Number of Parameters and Speed for Various Models on BUSI and CT2US Datasets

Model	**CT2US** ↑	**OpenKidney** ↑	**BUSI** ↑	**Params** ↓	**Speed (FPS)** ↑
SegResNet	0.977	**0.889**	0.799	**0.4**M	**21**
MobileNetV3	**0.979**	0.833	**0.810**	3.2M	10
ResUNet	0.975	0.877	0.765	13.3M	9
MobileViTV2	**0.979**	0.868	0.808	13.3M	7
Segformer	0.975	0.870	0.741	3.7M	9

Additionally, visual results from the segmentation tasks are presented in the Figure 2, providing a qualitative assessment of models performance. These visualizations further corroborate the quantitative findings, showcasing the precision of segmentation achieved model. Nevertheless, there exist problem of domain adaptation as images from different ultrasonic devices might have some dissimilarities that affect the results of model.

In the absence of the data from device we have and the focus of our project on hypothesis testing, we ignore the differences and deployed the model trained on OpenKidney dataset[13] as it is the closest one to our experimental setup. Figure 3 shows how the models perform on data captured from Clarius device.

7 Limitations

While the proposed method offers promising results in ultrasonic image segmentation, it is not without limitations. Primarily, the project served as a proof of concept for idea testing, and comprehensive real-life deployment would require more extensive development. One significant limitation is the need for a larger and more diverse ultrasonic dataset to ensure better generalization of the models across various scenarios. Additionally, direct integration with ultrasonic devices through an API for live streaming data would be essential for practical use. Future work should also focus on developing a more robust application that offers not just binary segmentation but also detailed organ part delineation and disease detection features, providing a more holistic diagnostic tool.

8 Conclusion

In conclusion, this project successfully demonstrates the feasibility of using advanced deep learning models like SegResNet, MobileNetV3, ResUNet, MobileViTV2, and Segformer for ultrasonic image segmentation on mobile devices. Our experiments, leveraging datasets such as BUSI, CT2US, and Open Kidney, indicate promising results in terms of accuracy as measured by the Dice score. However, the scope of this project was limited to idea testing, and further work is necessary to transition from concept to clinical application. This includes acquiring more diverse data for model generalization, establishing live data integration through an API, and expanding application features to include detailed organ segmentation and disease detection.

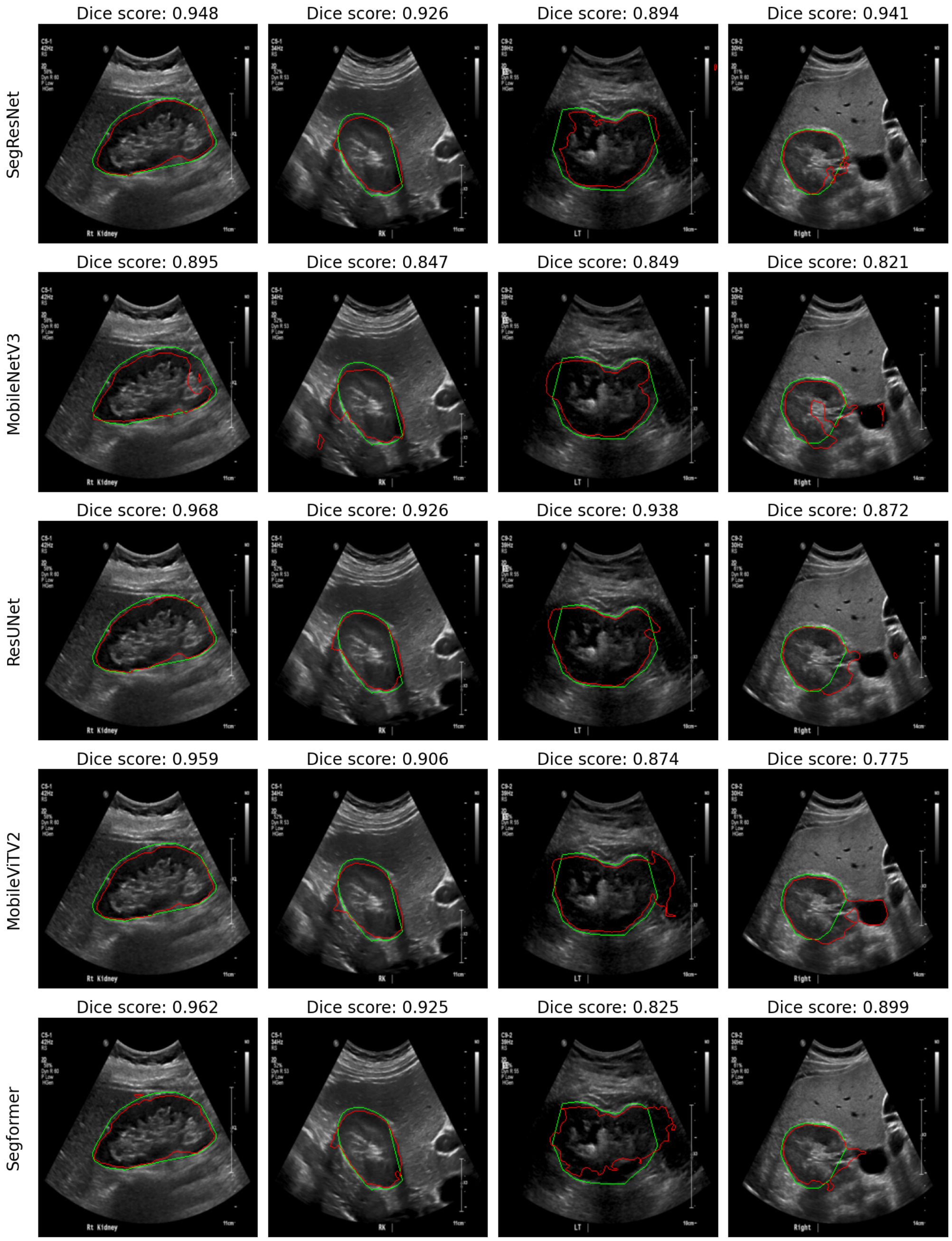


Figure 2: Examples of Ground Truth (Green) and Prediction (Red) on Open Kidney Dataset

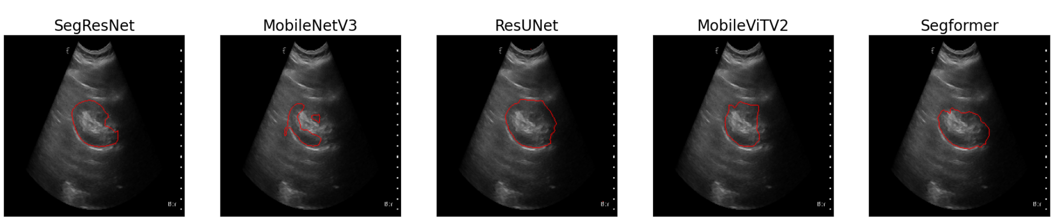


Figure 3: Examples of Prediction on Unseen Real Captured Data

References

[1] Clarius c3. https://clarius.com/scanners/c3/. Accessed: 2023-09-27.

[2] Pytorch mobile. https://pytorch.org/mobile/android/. Accessed: 2023-11-10.

[3] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in Brief*, 28:104863, 2020.

[4] Foivos I. Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:94–114, April 2020.

[5] Adrian Galdran, Gustavo Carneiro, and Miguel Ángel González Ballester. On the optimal combination of cross-entropy and soft dice losses for lesion segmentation with out-of-distribution robustness, 2022.

[6] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3, 2019.

[7] Ibrahem Kandel and Mauro Castelli. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express*, 6(4):312–315, 2020.

[8] Guoqi Liu, Jiajia Wang, Dong Liu, and Baofang Chang. A multiscale nonlocal feature extraction network for breast lesion segmentation in ultrasound images. *IEEE Transactions on Instrumentation and Measurement*, 72:1–12, 2023.

[9] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts, 2017.

[10] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019.

[11] Sachin Mehta and Mohammad Rastegari. Separable self-attention for mobile vision transformers, 2022.

[12] Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization, 2018.

[13] Rohit Singla, Cailin Ringstrom, Grace Hu, Victoria Lessoway, Janice Reid, Christopher Nguan, and Robert Rohling. The open kidney ultrasound data set, 2022.

[14] Yuxin Song, Jing Zheng, Long Lei, Zhipeng Ni, Baoliang Zhao, and Ying Hu. Ct2us: Cross-modal transfer learning for kidney segmentation in ultrasound images with synthesized data. *Ultrasonics*, 122:106706, 2022.

[15] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers, 2021.