International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 12, Issue 5 [May. 2023] PP: 40-51

AI Assisted Gastrointestinal Tract Segmentation for Radiation Therapy

Soorya Prakash K[1] , Vidhya S[2]
Department of Sensor and Biomedical Technology Vellore Institute of Technology, Vellore,India

ABSTRACT: *Gastrointestinal cancer is a prevalent disease worldwide, with approximately 6 million people diagnosed every year. Radiation therapy is a common treatment for this cancer ,in which X-ray beams are used to deliver high doses of radiation to tumors while avoiding the stomach and intestines.but the manual outlining of organs such as stomach and intestines is a time-consuming process for technicians, taking up to 50 minutes to 1 hour per patient. The implementation of an AI system that can assist this process will be beneficial to improve patient outcomes and reduce the workload on technicians. This research proposes the use of advanced deep learning architectures such as transformers and EfficientNet on top of U-Net Segmentation algorithm as encoders to improve the efficiency of outlining organs in MRI scans , making treatments faster and effective for patients.With this proposed assistive AI ,we can reduce the time technicians spend in outlining organs, enabling them to allocate more time to improving treatment for more patients .Detailed comparative analysis in this research has shown that using Le-VIT Transformer as U-Net Encoder significantly outperforms EfficientNet as U-Net Encoder in all aspects . We have achieved a Dice Coefficient 90.56% and Jaccard Score 80.73% with an increase of 12% in dice coefficient and 23% in Jaccard index compared to EfficientNet model ,indicating that these models are providing exceptional results and ready to be tested in the real world .*

Keywords: *Medical Image Segmentation, MRI Data, GastroIntestinal tract cancer, Organs segmentation , U-Net Segmentation , Le-VIT Transformer , Computer Vision , Efficient Net , Encoder-Decoder Architecture.*

Date of Submission: 25-04-2023 Date of acceptance: 05-05-2023

I. INTRODUCTION

Gastrointestinal (GI) cancer refers to a group of cancers that affect the digestive system, which includes the esophagus, stomach, liver, pancreas, small intestine, colon, rectum, and anus. These cancers can occur in any part of the digestive system and can spread to other parts of the body.[1]It accounts for approximately 25% of all cancer cases and deaths. According to the World Health Organization (WHO), approximately 4.5 million new cases of GI cancer are diagnosed each year, and it is the second leading cause of cancer-related deaths worldwide.[2] The most common types of GI cancer are colorectal cancer, stomach cancer, and liver cancer.

This cancer is a major health concern worldwide, with approximately 6 million people diagnosed with the disease in 2021[3] . Of these patients, half are eligible for radiation therapy, which typically lasts for 1-6 weeks. Radiation therapy is a commonly used treatment for gastrointestinal cancer, in which X-ray beams are used to deliver high doses of radiation to tumors while avoiding the stomach and intestines.However, the process of outlining tumors and intestines manually to adjust the direction of X-ray beams is time-consuming and labor-intensive, prolonging treatment time. With newer technology such as MR-Linacs, the daily position of tumors and intestines can be visualized, but the process of manually outlining them remains a bottleneck in treatment planning . On average, it takes around 50 minutes to 1 hour per patient to manually outline the stomach and intestines, which can be a significant burden on radiation oncologists [4].

One challenge in radiation therapy for GI cancer is ensuring that the radiation is delivered to the correct area and at the correct dosage. This is where AI assisted image segmentation can assist technicians in giving dosage in a streamlined way.By using medical imaging technology, such as CT scans or MRI scans, image segmentation can help technicians outline the organs and tissues that need to be targeted with radiation. [5]This helps ensure that the radiation is delivered precisely to the tumor and surrounding tissues, while avoiding healthy organs and tissues.

By improving the current U-Net architecture, we aim to better segment organs in images, which will reduce the time radiation oncologists take to outline organs. This may help automate the process, making treatments faster and more effective for more patients. With the help of advanced image segmentation techniques, radiation oncologists will be able to spend more time adjusting the doses, thereby improving the overall efficiency of treatment planning.

The objective of my paper is to explore the potential of advanced image segmentation techniques, specifically the combination of U-Net and transformers, to assist radiation oncologists in accurately and efficiently outlining organs at risk. By streamlining this process, we can reduce the time technicians spend outlining organs, enabling them to allocate more time to adjusting doses and improving treatment plans for more patients .

II.LITERATURE REVIEW

Image Segmentation has been researched in the medical field for a long time and with the invention of U-Net Segmentation in 2015 , it has boosted the application of AI in Medical Image Segmentation . In [6] authors have used ultrasound images of women with breast cancer between the age 25 and 75 and used an ensemble of Deep learning models such as CNN, MaskR CNN, U Net, and ResNet and have achieved 90% accuracy but detailed overview is yet to be discussed. In [7] authors have described a deep learning-based framework for medical image segmentation that aims to reduce the time and labor cost of label generation. The framework utilizes an auto-encoder to extract features from unlabeled images, applies feature clustering such as KNN for label generation,and trains a segmentation model in a semi-supervised manner. This is really a great idea to create labels as the quantity of data to be approved to use for AI is a main problem. In [8] We see an approach of using transformers and authors have proposed a network called MedSeq which addresses the limitation of current methods by exploiting the relationships between successive frames in medical image sequences and the dependencies within individual frames. The network is composed of two main parts: a Cross-frame Attention module, which learns correlations among frames, and a Boundary-aware Transformer, which improves the segmentation of boundary patches.

In [9] the authors present a novel method for simultaneously tracking centerlines and segmenting the small intestine in 3D cine-MRI scans. The method uses a stochastic tracker built on top of a CNN-based orientation classifier, and the segmentation is conditioned on the locations of the intestinal centerlines. They have used 3D cine-MRI scans. But these Images cannot be used for our problem as the data we have are 2D.In [10] the authors have discussed segmenting and classifying diseases in the gastrointestinal tract using WCE images. The images are preprocessed using filtering and contrast enhancement techniques and then segmented using a modified U-Net. From the segmented images, features are extracted and then given to an improved DNN for classification . In [11] In this paper they have worked polyps segmentation in the gastrointestinal tract using deep ensemble learning with a bagging based U-Net architecture (BaggedUNet). The proposed method trains several lighter U-Net architectures and combines their decisions using majority voting.they achieve 3%-9% improvement on different evaluation metrics. These are the 2 papers where we see the use of U-Net Architectures and both conclude that U-Net segmentation provides great improvement in the performance compared to the traditional methods.In [12] authors have proposed a method of generating new, plausible samples of gastrointestinal images using image-to-image translation models like pix2pix and Cyclegans.They have used 260 gastrointestinal images with size (512,512). This can be useful for us to generate more images for training . but the helpfulness of these images for the train model to work on real time data has to be analyzed. In [13] authors have worked on a method for multi-organ segmentation in CT and MRI images called MPSHT (Multiple Progressive Sampling Hybrid Model Multi-Organ Segmentation).

III.DATA

The data used in this thesis comprises MRI scan images of the abdomen captured from the top view.

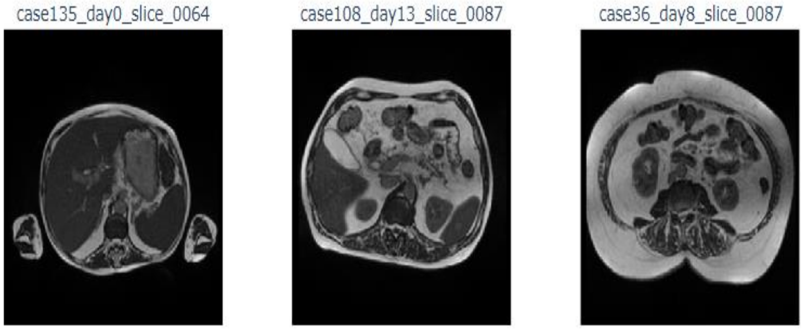


Fig 1 : Samples extracted from data

This data set contains a total of around 50000 scan images in 16-bit grayscale PNG format. The MRI

scans are from actual gastrointestinal cancer patients who had 1-5 MRI scans on separate days during their radiation treatment. The source of this data is from the UW-Madison Carbone Cancer Center, which is a renowned pioneer in MR-Linac based radiotherapy. The data is therefore of high quality and is well-suited for the development of deep learning models to assist in the segmentation of gastrointestinal organs, and to help optimize radiation treatment plans for cancer patients.
Source of the Data Link : [14]

Number of Samples per Organs in Data:

Dataset contains segmentation masks of three organs: large bowel, small bowel, stomach. These are all part of the digestive system. The bowels (small and large intestine) are responsible for breaking down food and absorbing the nutrients.

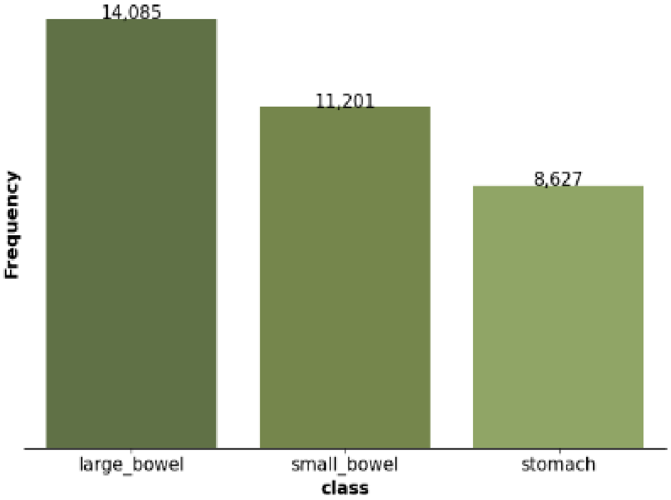


Fig 2 : Sample Distribution

The dataset contains 14,085 segmentations of Large Bowel (Large Intestine) , 11,201 segmentations of small intestine and 8,627 instances of Stomach.

Multi Label Classification and Dependent Variable Creation :

Multi-label classification is a machine learning task where each sample can be assigned to multiple classes simultaneously. In other words, instead of having a single target label for each instance, there can be multiple labels associated with each instance.Since We will be segmenting 3 organs in the same image , This becomes a multi label classification problem where we predict the probability for 3 organs in each pixel and use this data to create a mask. The Dependent variable / label will be a multi dimensional array with values (0 or 1) denoting a mask of whether an organ is present or not.

Now, for each image , we are going to create an image of shape [img height, img width, 3], where 3 (number of channels) are the 3 layers for each class:

- the first layer: large intestine
- the second layer: small intestine
- the third layer: stomach

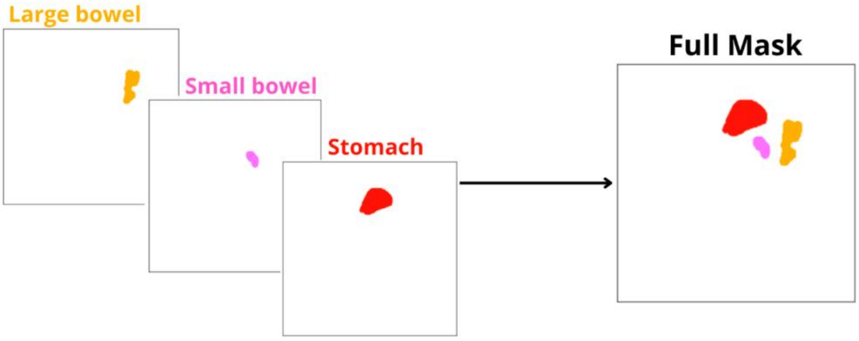


Fig 3 : Multi Dimensional mask Generation

Model Architectures used :

U-Net Segmentation Algorithm :

U-Net is a convolutional neural network architecture used for image segmentation tasks. It was proposed by researchers from the Computer Science Department at the University of Freiburg in 2015.[15] The name "U-Net" comes from the U-shape of the network architecture.U-Net was initially developed for biomedical image segmentation, specifically for segmenting neuronal structures in electron microscopy images. The architecture was designed to handle small training sets and to produce accurate segmentations with high spatial resolution.

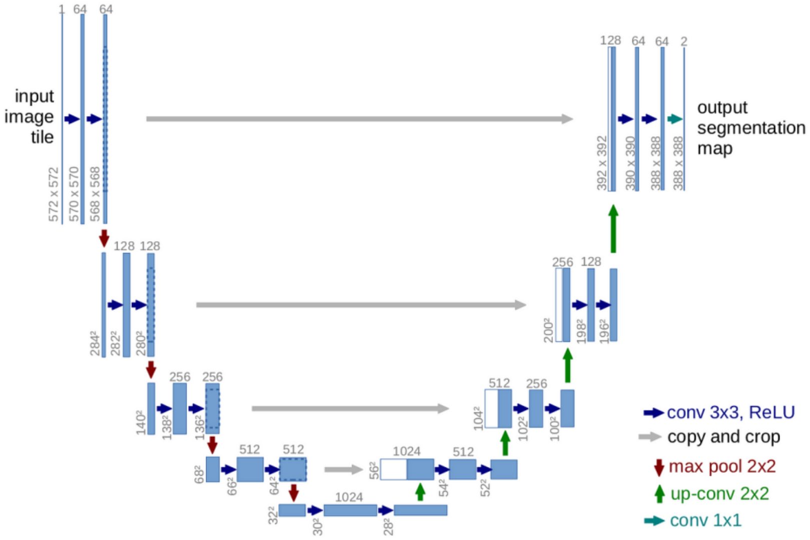


Fig 4 : U-Net Architecture

The U-Net architecture consists of a contracting path and an expanding path, which are connected by a bridge. The contracting path consists of convolutional and max-pooling layers, which reduce the spatial resolution of the input image while increasing the number of feature maps. The expanding path consists of transposed convolutional and concatenation layers, which increase the spatial resolution of the feature maps while reducing their number. The bridge consists of a bottleneck layer, which maintains a high level of spatial resolution while also capturing contextual information.[16]

Efficient Net Algorithm :

EfficientNet is a convolutional neural network architecture that was proposed by researchers at Google in 2019.[17] It is designed to achieve state-of-the-art performance with fewer parameters and FLOPS compared to other popular architectures like ResNet and Inception.EfficientNet achieves this efficiency by using a combination of techniques like compound scaling, which involves scaling the network's width, depth, and resolution simultaneously, and using a novel mobile inverted bottleneck block.

EfficientNet as UNet Architecture :

EfficientNet can be used as an encoder in the U-Net architecture for image segmentation tasks. The U-Net architecture is a popular and effective approach for image segmentation, especially in medical imaging.

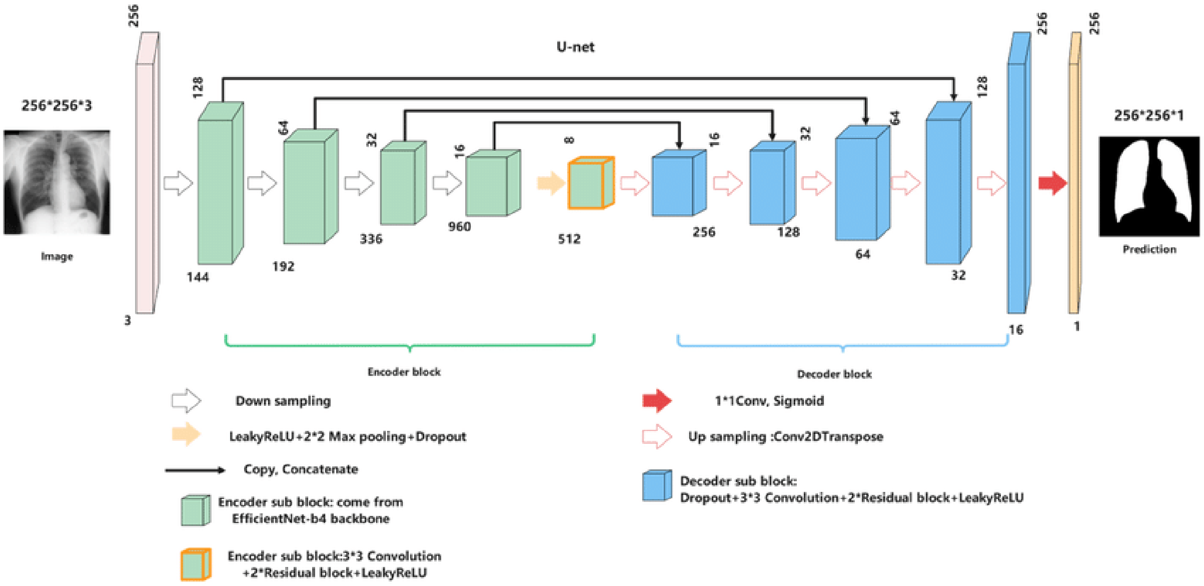

U-net
256*256*3
Image
256
3
128
144
64
192
32
336
16
960
8
512
16
256
32
128
64
64
128
32
256
16
256
1
256*256*1
Prediction
Encoder block
Decoder block
Down sampling
LeakyReLU+2*2 Max pooling+Dropout
Copy, Concatenate
Encoder sub block: come from EfficientNet-b4 backbone
Encoder sub block:3*3 Convolution +2*Residual block+LeakyReLU
1*1Conv, Sigmoid
Up sampling :Conv2DTranspose
Decoder sub block: Dropout+3*3 Convolution+2*Residual block+LeakyReLU

Fig 5 : EfficientNet as UNet Encoder Architecture

Using EfficientNet as the encoder in the U-Net architecture can benefit the segmentation task in several ways. [18]Firstly, EfficientNet is a highly efficient and powerful convolutional neural network architecture, which can extract rich features from the input image. Secondly, the use of EfficientNet can help to improve the accuracy and efficiency of the segmentation task.To use EfficientNet as the encoder in the U-Net architecture, the pre-trained EfficientNet model is first loaded and its weights are frozen. The input image is then passed through the EfficientNet model, which extracts high-level feature maps at multiple scales. These feature maps are then fed into the decoder part of the U-Net, which produces the final segmentation map.

LeVIT Transformers as Encoder in U-Net Architecture :

LeViT (LeViT-128 and LeViT-256) is a recent transformer-based model that was designed specifically for image classification tasks.[19] LeViT uses a hybrid architecture that combines convolutional and transformer layers, enabling it to process images more efficiently than traditional transformer-based models.The LeViT architecture consists of three stages. In the first stage, the input image is processed by a set of convolutional layers, which extract low-level features. In the second stage, the features are processed by a set of transformer blocks, which enable the model to learn long-range dependencies. Finally, in the third stage, the features are processed by another set of convolutional layers, which enable the model to produce the final output.[20]

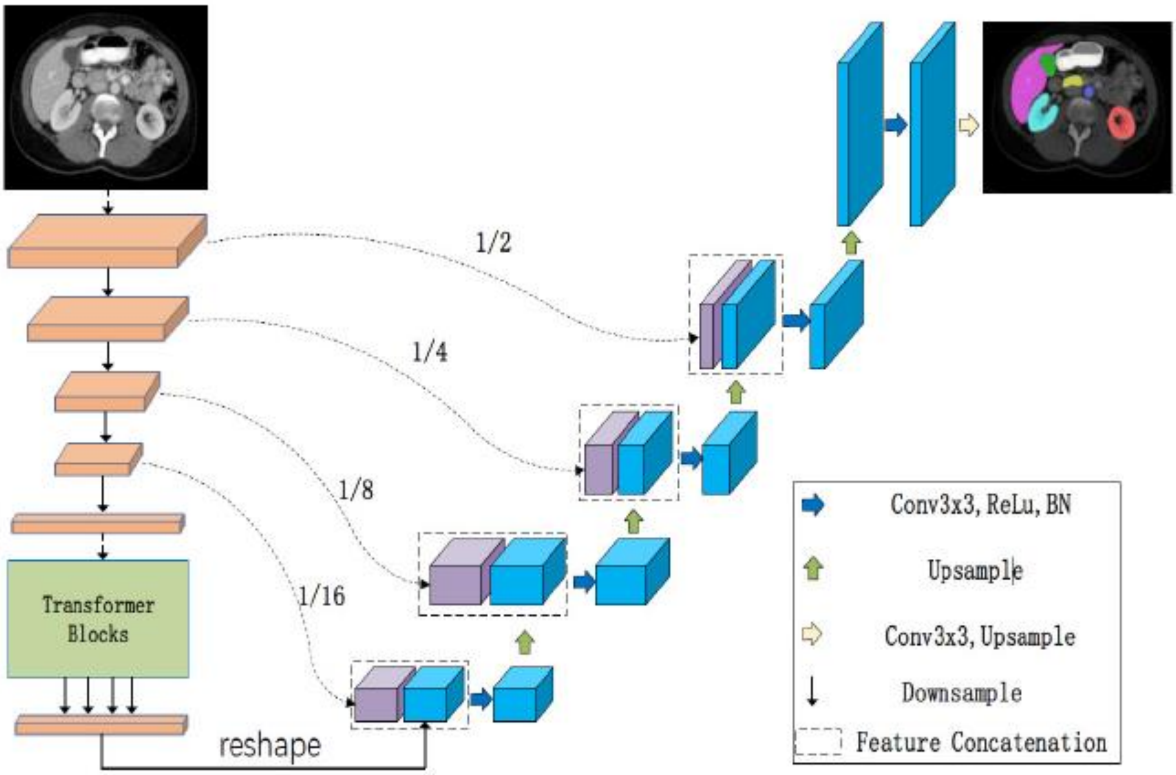


Fig 6 : LeVIT Transformer as Encoder in U-Net Architecture

LeViT can also be used as an encoder in the U-Net architecture for image segmentation tasks. In this approach, the LeViT model is pre-trained on a large-scale image classification dataset, such as ImageNet. Then, the pre-trained LeViT model is used as the encoder in the U-Net architecture. [21] The LeViT encoder extracts high-level feature maps from the input image, which are then fed into the decoder part of the U-Net to produce the final segmentation map.Using LeViT as the encoder in the U-Net architecture for image segmentation tasks can provide several benefits. Firstly, LeViT is a highly efficient and accurate model, which can extract rich feature representations from the input image. Secondly, the use of LeViT can help to improve the accuracy and speed of the segmentation task. Finally, this approach has shown promising results in various applications, including medical image segmentation.

Constraints :

The biggest constraint while solving this problem is the lack of computationally heavy resources such as TPUs for training . But thinking about the problem in the production level , I took this as a challenge and created a model which is not computationally expensive so that this can be used on light devices installed in any hospitals.

Metrics to Quantify the Performance :

Dice Coefficient :

Dice coefficient, also known as F1 score, is a commonly used metric for evaluating the performance of image segmentation algorithms. The Dice coefficient measures the overlap between the predicted segmentation mask and the ground truth segmentation mask.[22]The Dice coefficient is defined as twice the intersection of the predicted mask and the ground truth mask, divided by the sum of the areas of the two masks.

Jaccard Score :

The Jaccard score, also known as the Intersection over Union (IoU) score, is a commonly used metric for evaluating the performance of image segmentation algorithms.[23] It measures the similarity between the predicted segmentation mask and the ground truth segmentation mask.

The Jaccard score is defined as the ratio of the intersection of the predicted mask and the ground truth mask to the union of the two masks.

Loss Functions to Train the Model :

Since our problem is Multi Label Classification , loss functions are chosen in a way that they optimize this kind of problem .

Soft Binary Cross Entropy :

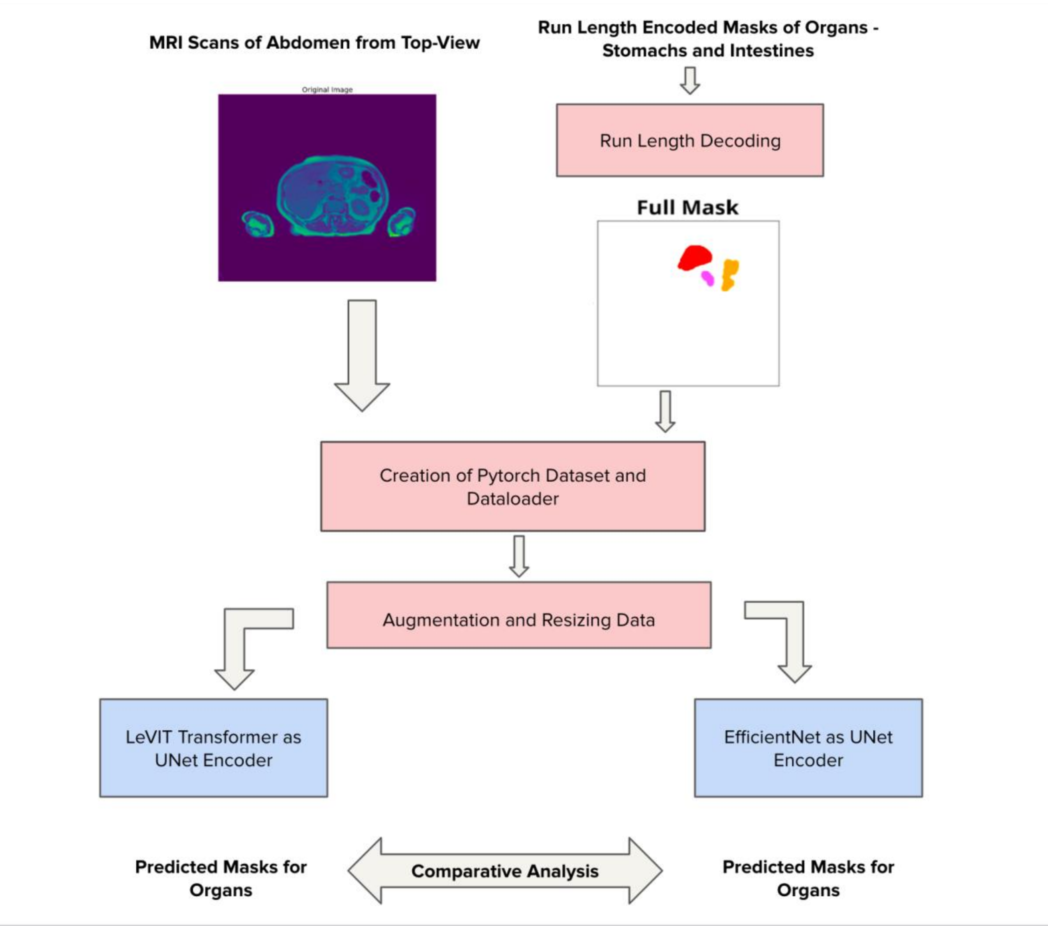
Soft binary cross entropy is a loss function used in image segmentation tasks that involve multi-label classification.[24] In traditional binary cross entropy, the predicted output is compared with a binary ground truth, where each pixel can only belong to one class or background. However, in some cases, each pixel may belong to multiple classes, which makes it difficult to use binary cross entropy as the loss function.

Tversky Loss :

Tversky loss is a loss function used in image segmentation tasks that involve multi-label classification. It is an extension of the dice loss function and is designed to handle imbalanced datasets, where one or more classes may be underrepresented.

III.METHODOLOGY

The Goal of the project is to establish an Assistive AI which can help radiation technicians outline the organs such as Stomachs and Intestines faster so that we can reduce the overall time consumption of organs from one hour to 10-15 minutes.



MRI Scans of Abdomen from Top-View
Original Image
Run Length Encoded Masks of Organs - Stomachs and Intestines
Run Length Decoding
Full Mask
Creation of Pytorch Dataset and Dataloader
Augmentation and Resizing Data
LeVIT Transformer as UNet Encoder
EfficientNet as UNet Encoder
Predicted Masks for Organs
Comparative Analysis
Predicted Masks for Organs

Fig 7: High Level View of the approach

The data used in this thesis comprises MRI scan images of the abdomen captured from the top view and it is from UW Madison Research Centre . Masks for each organ are run length encoded in the dataset determining the pixels of each organ . These run length encoded strings will be decoded to identify the location of organs and a multidimensional array of masks are created to be given as a label to our deep learning model. Since Deep Learning really tends to perform well with large amounts of data , augmentations are done to increase the number of samples .

The training data is used to train two model architectures : LeVisual Transformer as UNet Encoder and EfficientNet as UNet Encoder which trains on the given data and predicts an multi dimensional array of binary values , each channel representing one organ and binary values (1 -denoting the presence of organ , 0- Not) helping us create masks which better outline the organs and help assist the technicians .

Comparative Analysis is done on different aspects of these models such as the performance in different metrics , the validation loss and the time and power consumption to generate these models to get a detailed understanding of how effective the model will perform in real time.

IV.RESULTS AND DISCUSSION

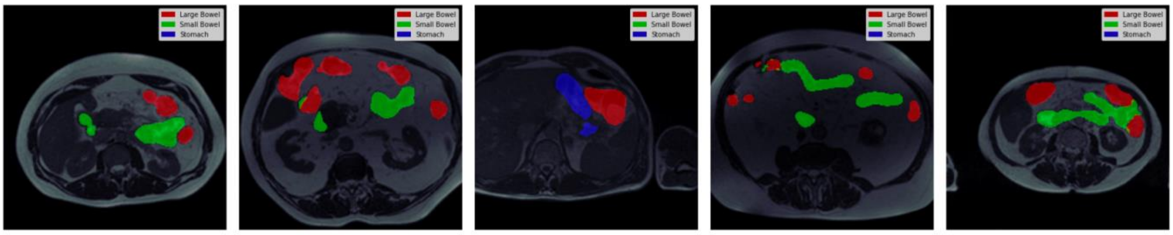


Fig 8: Le-VIT Transformer model predicted segmentation

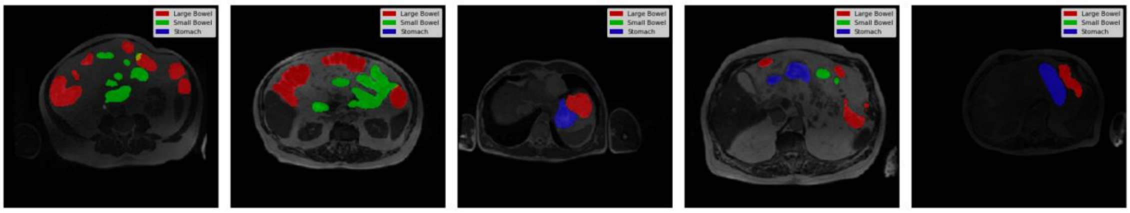


Fig 9: EfficientNet model predicted segmentation

The Predicted masks by Le-VIT Transformer as U-Net Encoder are generally better in comparison to Efficient Net As U-Net Encoder visually , random samples were taken and visual comparison was made to make an initial judgment . Metrics and Loss function comparison is done to understand the difference better.

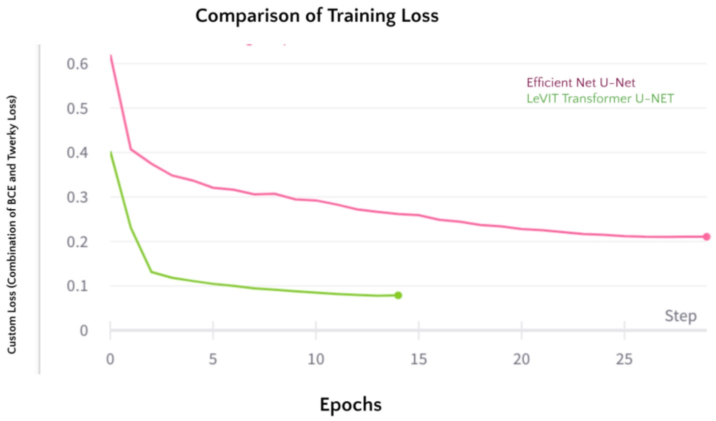


Fig 10: Comparison of Training loss

A Custom Loss of Equal weighted sum of Binary Cross Entropy and Twerky loss was used to train the model . LeVIT Transformer gave exceptional results by reaching 0.078 as its loss in just 14 epochs while even though the EfficentNet model was trained for 30 epochs its final loss value was 0.21 .As Training loss is not a best indicator for performance , the paper also discusses the Validation loss and the metrics .

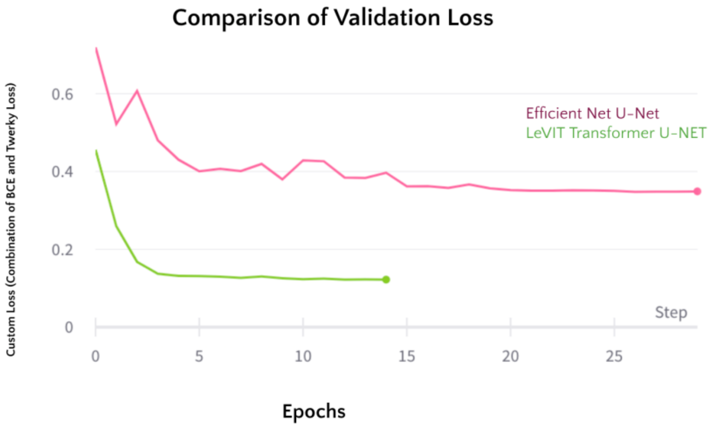


Fig 11 :Comparison of Validation Loss

LeVIT Transformer outperforms the Efficient Net model by a huge margin in the Validation loss too . This proves that the LeVIT performance is not due to overfitting and it is due to the model's better ability to learn with the help of Global representation Learning of Transformers. Validation loss is 0.1211 at its 14th epoch but even though the Efficient net model trained for 30 epochs , the minimal loss it could attain was 0.3419 .

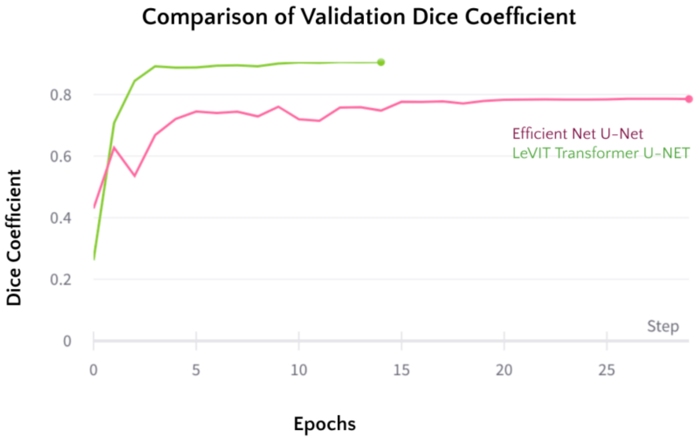


Fig 12 : Comparison of Dice Coefficient Metric

Dice coefficient is used for image segmentation to evaluate the similarity between the predicted and ground truth segmentations. It measures the overlap between the two sets of pixels, with higher values indicating better performance. LeVIT Transformer UNet outperforms the Efficient Net -UNet by 12% by reaching an impressive score of 90% in just 14 epochs compared to Efficient Net Performance of 78% . This proves that the LeVIT Transformers provide exceptional Segmentations.

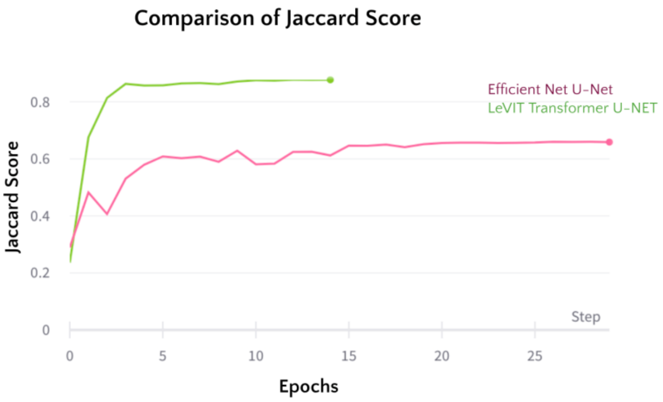


Fig 13 :Comparison of Jaccard Score

Jaccard score, also known as Intersection over Union (IoU), is a metric used for evaluating the accuracy of image segmentation models. It measures the overlap between the predicted and ground truth segmentations by dividing the intersection of the two sets of pixels by their union.LeVIT Transformer as U-Net Encoder achieves a score of 87.73% compared to the score of 65.8% of Efficient Net U-Net .

LeVIT Transformer as U-Net performs better over all metrics over Efficient Net Indicating its dominance in performance and its Better Segmentation Capabilities.Comparison of Scores gives us only one aspect of performance ,so We also compare the GPU usage of both the models to compare its efficiency and its carbon footprint to train the model.

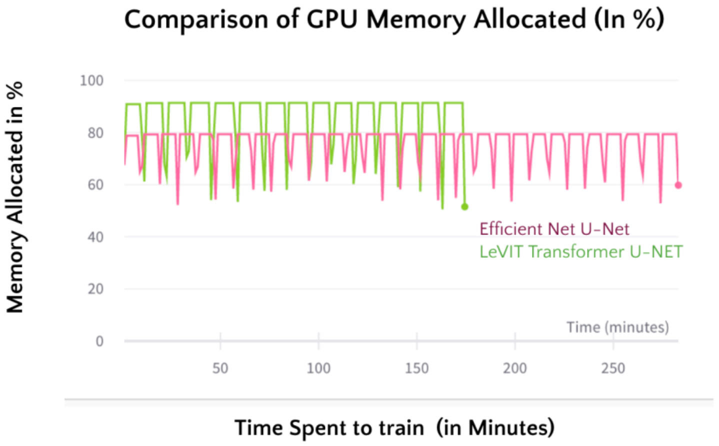


Fig 14 : Comparison of GPU Memory Allocation :

The % GPU Memory Usage is the only case where LeVIT UNet is less efficient than Efficient UNet . The LeVIT UNet Model uses 90% of memory during its training process compared to 80% memory usage of Efficient Net UNet Architecture.

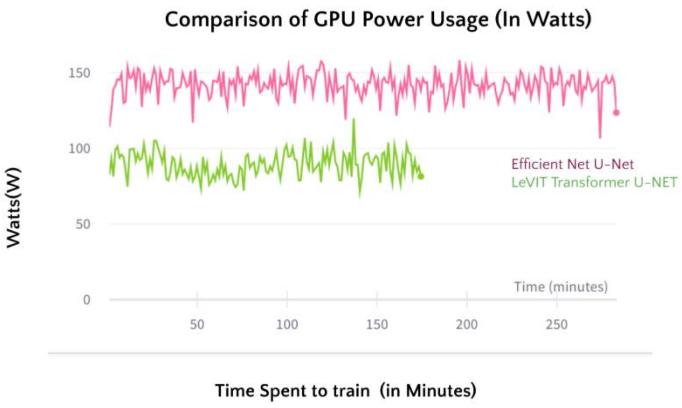


Fig 15 : Comparison of GPU Power Usage :

The GPU power usage of Efficient Net - UNet was 154 watts compared to the peak usage of 119 watts of Le-VIT Transformers . We can also notice that the average GPU power usage is around 93 watts for Le-VIT transformers as U-Net Encoder compared to 143 Watts of Efficient Net as U-Net. This also helps us realize that LeVIT achieves better scores by using only less power . This is a great advantage for the model .

V.CONCLUSION

The detailed comparative analysis in this research has proved our hypothesis that using Le-VIT Transformers as U-Net Encoder have great improvement in improvement compared to using EfficientNet as U-Net Encoder. We can see that the Transformers outperforms the EfficientNet in every aspect such as Dice Coefficient , Jaccard Index by large margin while consuming less power to train .

Comparative Analysis of Le-VIT Transformer and EfficientNet as U-Net Encoder

Metric	Le-VIT Transformer as U-Net Encoder	EfficientNet as U-Net Encoder	Information on Metric
Visual Comparison	Better	Poor -Average	
Training Loss	0.078	0.21	Lower the Better
Validation Loss	0.1211	0.3419	Lower the Better
Dice Coefficient	90.56%	78.55%	Higher the Better
Jaccard Score	87.73%	65.87%	Higher the Better
Peak GPU Memory Usage (%)	90%	80%	Lower the Better
Average GPU Power Usage (Watts)	93	143	Lower the Better

These exceptional results signify that using the Le-VIT transformer can assist Technicians to outline organs such as the stomach and intestines effectively , reducing the delay in improvement and allowing technicians to give better treatment.

REFERENCES

[1]. Types of Gastrointestinal Cancer | Penn Medicine

[2]. Global burden of 5 major types of gastrointestinal cancer – IARC

[3]. Cancer Statistics : https://www.cancer.org/cancer/gastrointestinal-stromal-tumor.html

[4]. Radiation Therapy

[5]. CT Scan (Computed Tomography): What is It, Preparation & Test Details

[6]. D. Yuvaraj, A. K. Sampath, N. Shanmugapriya, B. Samatha, S. Arun and R. Thiyagarajan : Medical Image Segmentation of Biomedical Images with Deep Convolutional Neural Networks using Ensemble Approach ,2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1041-1049, doi: 10.1109/ICOSEC54921.2022.9952142.

[7]. K. Huang, J. Huang, W. Wang, M. Xu and F. Liu : A Deep Active Learning Framework with Information Guided Label Generation for Medical Image Segmentation ,2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Las Vegas, NV, USA, 2022, pp. 1562-1567, doi: 10.1109/BIBM55620.2022.9995046.

[8]. Y. Zheng, B. Wang and Q. Hong, "UGAN: Semi-supervised Medical Image Segmentation Using Generative Adversarial Network," 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 2022, pp. 1-6, doi: 10.1109/CISP-BMEI56279.2022.9980009.

[9]. Louis D. van Harten, Catharina S. de Jonge, Kim J. Beek, Jaap Stoker, Ivana Išgum, Untangling and segmenting the small intestine in 3D cine-MRI using deep learning,Medical Image Analysis,Volume 78,2022,102386,ISSN 1361-8415, https://doi.org/10.1016/j.media.2022.102386.

[10]. Vrushali Raut, Reena Gunjan, Virendra V. Shete & Upasani Dhananjay Eknath (2022) Gastrointestinal tract disease segmentation and classification in wireless capsule endoscopy using intelligent deep learning model, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, DOI: 10.1080/21681163.2022.2099298

[11]. S. M. Faraz Ali, M. A. Tahir and A. B. Khalid, "BaggedUNet: Deep Machine Vision approach for Polyps Segmentation in Gastrointestinal Tract," 2022 24th International Multitopic Conference (INMIC), Islamabad, Pakistan, 2022, pp. 1-7, doi:10.1109/INMIC56986.2022.9972945.

[12]. Adjei, Prince & Lonseko, Zenebe & Rao, Nini. (2020). GAN-Based Synthetic Gastrointestinal Image Generation. 338-342. 10.1109/ICCWAMTIP51612.2020.9317341.

[13]. Zhao Y, Li J, Hua Z. MPSHT: Multiple Progressive Sampling Hybrid Model Multi-Organ Segmentation. IEEE J Transl Eng Health Med. 2022 Sep 26;10:1800909. doi: 10.1109/JTEHM.2022.3210047. PMID: 36457896; PMCID: PMC9704745.

[14]. UW-Madison GI Tract Image Segmentation | Kaggle , UWCCC Research – Resource for Researchers – UW–Madison (wisc.edu)

[15]. Ronneberger, Olaf & Fischer, Philipp & Brox, Thomas. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. LNCS. 9351. 234-241. 10.1007/978-3-319-24574-4_28.

[16]. https://developers.arcgis.com/python/guide/how-unet-works/

[17]. Tan, Mingxing & Le, Quoc. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.

[18]. B. Baheti, S. Innani, S. Gajre and S. Talbar, "Eff-UNet: A Novel Architecture for Semantic Segmentation in Unstructured Environment," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 2020, pp. 1473-1481, doi: 10.1109/CVPRW50498.2020.00187.

[19]. Graham, Benjamin and El-Nouby, Alaaeldin and Touvron, Hugo and Stock, Pierre and Joulin, Armand and Jegou, Herve and Douze, Matthijs :LeViT: A Vision Transformer in ConvNet's Clothing for Faster Inference ,Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)},October,2021.

[20]. LeViT

[21]. Xu, Guoping and Zhang, Xuan and Fang, Yin and Cao, Xinyu and Liao, Wentao and He, Xinwei and Wu, Xinglong, LeVit-UNet: Make Faster Encoders with Transformer for Biomedical Image Segmentation. http://dx.doi.org/10.2139/ssrn.4116174

[22]. Sørensen–Dice coefficient - Wikipedia

[23]. Jaccard Index / Similarity Coefficient - Statistics How To

[24]. Understanding binary cross-entropy / log loss: a visual explanation | by Daniel Godoy | Towards Data Science

Image Credits :

[Fig 4] UNet — Line by Line Explanation. Example UNet Implementation | by Jeremy Zhang | Towards Data Science

[Fig 5]]https://www.researchgate.net/figure/The-architecture-of-U-Net-with-EfficientNet-b4-Encoder_fig2_360804936

[Fig 6] https://sh-tsang.medium.com/review-levit-unet-make-faster-encoders-with-transformer-for-medical-image-segmentation-a5ce4ad22581