GI Tract Image Segmentation

1st Md.Abu Obaida Ma-az
Department of EEE
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
1906007@eee.buet.ac.bd

2nd Tapu Dutta
Department of EEE
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
1906049@eee.buet.ac.bd

3rd A.S.Al Mahmud Sajid
Department of EEE
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
1906050@eee.buet.ac.bd

4th Md.Ramim Hasan Shawn
Department of EEE
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
1906082@eee.buet.ac.bd

***Abstract*—Accurate and automated segmentation of gastrointestinal (GI) cancer images is crucial during radiation therapy. The goal is to precisely target the tumor with X-ray beams while sparing healthy regions of the stomach and intestines. Therefore, there is a pressing need for precise GI image segmentation in clinical settings.Unet and Unet++ was proposed in this paper for gastrointestinal image segmentation in CT and MRI. This project combines CNN and transformer models with a weighted module and achieved high Dice score and Jaccard score on segmentation datasets.This method can be extended to medical segmentation in different modalities and can provide decision support for clinical radiotherapy plans.**

I. Introduction

Human gastrointestinal (GI) tract is important and susceptible to infections. GI tract infections cause 1.8 million deaths annually. Gastrointestinal cancer is a prevalent and dangerous malignancy. Radiation therapy can improve cure rates in GI cancer patients. The objective is to improve dose delivery to the tumor while avoiding the stomach and intestines.The mother paper focuses on the need for accurate and automated gastrointestinal image segmentation in clinical practice. It introduces two medical image segmentation framework called Unet and Unet++. Unet++ combines the strengths of CNN and transformer models for global and local feature fusion.

The proposed method achieves superior segmentation performance in gastrointestinal segmentation tasks. It can be easily extended to medical segmentation in different modalities such as CT and MRI. The method provides decision support for clinical radiotherapy plans.An ML strategy was introduced in the network to supervise the decoder's learning process at different scales and improve the quality of the generated features. Different deep learning models and the proposed Unet++ model was evaluated on the UWMGI and Synapse datasets and the model demonstrated advanced segmentation performance. Medical image segmentation serves as the cornerstone for computer-aided diagnosis and treatment planning, and also finds its application in GI tract analysis. The proper outlining of GI tract image structures, such as the mucosal and submucosal layers, is very essential in diagnosis of diseases like ulcerative colitis, Crohn's disease, and gastrointestinal cancers. The manual segmentation by expert clinicians is still the golden standard yet it is time-consuming and error-prone.

In the last years, deep learning-based segmentation approaches have proved to be outstanding methods for the automation and increasing the precision of medical image segmentation procedures. Needless to say, the U-Net and its variants stand out among all approaches with impressive achievements in medical imaging.

The U-net model, which was put forward by Ronneberger et al. in 2015, is a convolutional neural network (CNN) tailored to the needs of biomedical image segmentation. The crescent shape, where it contracts to define the context and expands to yield better localization accuracy, has made it a sought-after technology in medical image analysis tasks.

The U-Net++ architecture developed by Zhou et al. is an extension of U-Net which includes dense connections and deep supervisions mechanisms. U-Net++ introduces feature reuse and propagation and takes care of the vanishing gradient problem. As a result, it gives a better segmentation performance.

In this paper, efficacy in GI tract image segmentation by means of U-Net and U-Net++ architectures across various datasets is investigated. We examine the fidelity of their performance in delineating regions of interest in endoscopic and histo-pathological imaging of the gastrointestinal tract.

II. Methodology

A. Dataset Description

The dataset used in this work was collected from Kaggle, provided by UW-Madison. The primary dataset contains 38.5 thousand grayscale PNG image data, and images were taken from MRI scans. The dataset contains scanned images of 85 cases, each containing scanned images of different days. Furthermore, RLE-encoded masks were employed as annotations for loss-less data compression. Another dataset different from the competition dataset was also used to run this experiment.

B. Data Processing & Augmentation

Because machine learning is a data-driven technique, data pre-processing is a crucial machine-learning task. Data that is sloppy or unprepared produce inefficient and incorrect results. Therefore data preparation is critical. Additionally, by using other operations like rotation, cropping,resizing, contrast,flipping, hue etc. are used to create new dataset from existing dataset. The resolution of all images in the dataset was not the same. As a result, the image size was converted to 224 x 224. Image data were masked based on three different output classes: the stomach, the large bowel, and the small bowel. In mask data, segmented areas of the healthy stomach, the small bowel, and the large bowel were represented in the blue, green, and red areas, respectively. Some images were augmented by rotating and changing contrast to bring variation.

1) Rotation: Rotation affects the general orientation of the image by the target angle. It can move here-and-there clockwise or anti-clockwise. People often make other edits to their photos such as rotating them, fixing their alignment, or making some art work adjustments.

2) Cropping: Crop refers to removing portions of the picture that are not important in order to bring out the focus of the image. It helps remove different distractions and contributes to obtaining the perfect shot. Often, the crop is carried out to remove the unnecessary background or to resize the pictures in certain ratios.

3) Flipping: Inverting an image horizontally or vertically is what flipping is about. Vertical flipping flashes the image along the vertical axis, whereas horizontal flipping mirrors it across the horizontal axis. Flipping is commonly applied for aesthetic purposes such as symmetry, or turning images upside.

4) Resizing: Picture resizing includes either maintaining aspect ratio or not scaling it another way. An image can be resized up or down. It is often employed to fix an image to a fixed space, compress files or prepare them to be displayed with different screen resolutions.

III. Proposed Architecture

A. Unet Architecture

The U-Net architecture proposed by Ronneberger et al. in 2015 has become a milestone in medical image segmentation. From its core, U-Net is a symmetric encoder-decoder network architecture designed for capturing both local and global contextual information and maintaining spatial resolution. Such a U-shaped architecture affordably detects object boundaries, which is widely applicable to biomedical image segmentation tasks.

1) Encoder:: The encoder part of U-Net includes a set of convolutional and pooling layers structured as a contracting pathway. This process is responsible for capturing hierarchical features at different scales in which the network can learn to obtain abstract representations of the input image. Every convolutional layer is surpassed by ReLU (rectified linear unit) activation function which enables to filter nonlinear feature extraction. As the encoder's processing goes on, the number of feature maps decreases and the number of feature channels increases, thus allowing the network to sequentially abstract and summarize information from the image. The encoder's contracting pathway ends in a bottleneck layer, which maximizes the net's receptive field and facilitates the extraction of semantic high-level features.

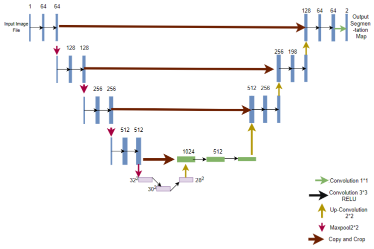


Fig. 1: Unet Architecture

2) Decoder:: The decoder component of the U-Net is symmetrical to the encoder's architecture. This process uses the pathway that brings back the spatial information that was lost during encoding and produces segmented masks with pixel-wise precision. Each layer in the decoder mixes upsampling and convolutional processes which are responsible for the restoration of high-resolution feature maps. Residual blocks, stimulated by the concept of residual learning, use the same layers of the encoder and decoder paths. Such skip connections facilitate the gradual blending between low-level and high-level features, hence enhancing the segmentation capability of the network.

3) Final Output Layer:: In the final level of the U-Net architecture, the convolutional layer uses the softmax activation function to compute.When this setup is in place, the marking zone for every pixel in the input image is calculated.Probability of each pixel is computed taking into account class (or object category) so that pixels that have higher probabilities while the ones having a lower probability are the ones that are accurately segmented.

Fig.1 comprises the encoder and decoder part of Unet Model in a single picture where skip connections can also be seen.

B. Unet++ Architecture

The U-Net++ architecture, first put forth by Zhou et al., was a breakthrough in the field of semantic segmentation, especially in the medical imaging domain. Complementing the core ideas of the U-Net structure, notably feature extraction, context integration, and segmentation accuracy improvement.

1) Encoder:: In the U-Net++ architecture, the encoder component is the component that extracts hierarchical features at different levels from the input image. The same as the original U-Net, the encoder consists of the convolutional layers that are arranged in the contracting pathway. In the convolutional path, the layers gradually downsample the input image, extracting high-level semantic features while simultaneously decreasing spatial size. Dense links at each resolution level create holistic feature extraction, letting the network to grasp complex spatial relationships and fine details of images

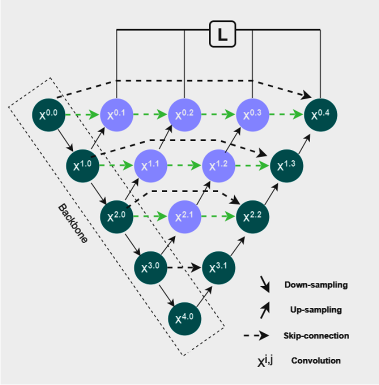


Fig. 2: Unet++ Architecture

simultaneously. Encoder's hierarchical characteristic of feature extraction serves as the base for subsequent steps of decoding.

2) Decoder:: It is the decoder part of U-Net++ that creates high-resolution feature maps from the abstraction generated by the encoder. It is a symmetric design with an encoder that is a double, consisting of a series of upsampling and convolutional layers arranged in a wide way. Skip connections, borrowed from the idea of residual learning, join corresponding layers from the encoder and decoder paths, thereby allowing the smooth flow of low-level and high-level features. It allows for image localization and segmentation with high precision. Furthermore, deep supervision mechanisms take place within the decoder and upgrade feature refinement as well as the integration of multi-scale contextual information which in turn improve the segmentation accuracy as well as the robustness.

3) Final Output Layer:: The final stage of the U-Net++ architecture is with a convolutional layer having a softmax activation function. It forms a probability map over the whole input image, and each pixel's probability shows the next extent to which it belongs to a specific category or object class. The softmax activation function serves an important role here because of the property that ensures the probability of all classes shall equal 1, which is useful for pixel-wise segmentation. The last output layer is embedded with network segmentation predictions hence enables articulate segmentation of objects and structures from the given image.

The U-Net++ is an upgraded version of the U-Net architecture with certain amendments.

Nested Skip Connections: U-Net++ architecture uses the nested skip connections of a network to aggregate the features while decoding the encoded data. Through gathering these features from various routes within the network, the model will achieve a higher segmentation mask accuracy. Deep Supervision: The U-Net++ architecture utilizes the deep supervision to enrich the model's performance through the network regularization during the training process. Fig.2 comprises the encoder and decoder part in a single picture where a skip connection can also be seen. By bridging the semantic gap between the encoder and decoder feature map, the Unet++ model is formed.

IV. Results

A. Whole Dataset

The model was ran on the whole masked dataset found on kaggle competition and the run history and the run summary found are showed in Fig.3 and Fig.4.

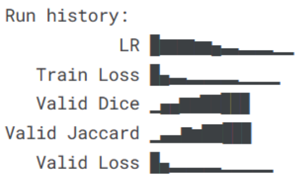
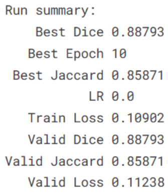


Fig. 3: Run History



```
Run summary:
    Best Dice 0.88793
   Best Epoch 10
 Best Jaccard 0.85871
           LR 0.0
   Train Loss 0.10902
   Valid Dice 0.88793
Valid Jaccard 0.85871
   Valid Loss 0.11238
```

Fig. 4: Run Summary

B. Large Bowel Dataset

The model was ran on the large bowel dataset extracted from the whole masked dataset found on kaggle competition and the run history and the run summary found are showed in Fig.5 and Fig.6.

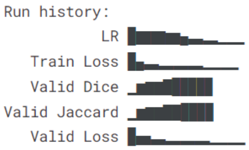
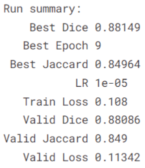


Fig. 5: Run History



```
Run summary:
    Best Dice 0.88149
   Best Epoch 9
 Best Jaccard 0.84964
           LR 1e-05
   Train Loss 0.108
   Valid Dice 0.88086
Valid Jaccard 0.849
   Valid Loss 0.11342
```

Fig. 6: Run Summary

C. Small Bowel Dataset

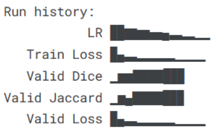
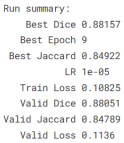


Fig. 7: Run History



```
Run summary:
    Best Dice 0.88157
   Best Epoch 9
 Best Jaccard 0.84922
           LR 1e-05
   Train Loss 0.10825
   Valid Dice 0.88051
Valid Jaccard 0.84789
   Valid Loss 0.1136
```

Fig. 8: Run Summary

The model was ran on the small bowel dataset extracted from the whole masked dataset found on kaggle competition and the run history and the run summary found are showed in Fig.7 and Fig.8.

D. Stomach Dataset

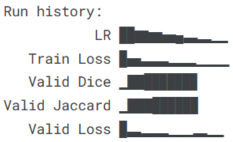
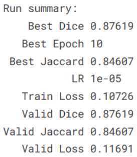


Fig. 9: Run History



```
Run summary:
    Best Dice 0.87619
   Best Epoch 10
 Best Jaccard 0.84607
           LR 1e-05
   Train Loss 0.10726
   Valid Dice 0.87619
Valid Jaccard 0.84607
   Valid Loss 0.11691
```

Fig. 10: Run Summary

The model was ran on the stomach dataset extracted from the whole masked dataset found on kaggle competition and the run history and the run summary found are showed in Fig.9 and Fig.10

V. Comparison

Model	Paper Scores		Our Scores	
	Dice Score	IoU Score	Dice Score	IoU Score
UNet	0.8826	0.8525	0.8818	0.8515
UNet++		0.8592	0.8879	0.8587

TABLE I: Results of UNet and UNet++

In TABLE I, the outcome of UNet and UNet++ model has been shown. The metrics for comparison are Dice Score and IoU Score. From the table, the experiment conducted by the authors of this paper was pretty similar with the mother paper. However, the metrics slightly improves in case of UNet++ model.

	WholeDataset	LargeBowel	SmallBowel	Stomach
Dice Score	0.88793	0.88149	0.88157	0.87619
Jaccard Value	0.85871	0.84964	0.84922	0.84607
Train Loss	0.10902	0.108	0.10825	0.10726
Valid Dice	0.88793	0.88086	0.88051	0.87619
Valid Jaccard	0.85871	0.849	0.84789	0.84607
Valid Loss	0.11238	0.11342	0.1136	0.11691

TABLE II: Comparison of Results between Different Datasets

In Table II, four cases and their corresponding outcomes have been shown. These four cases are, 'Whole Dataset', 'Large Bowel Dataset', 'Small Bowel Dataset' and 'Stomach Dataset'. Whole dataset was splitted into three different parts in accordance with each class: 'large bowel, 'small bowel' and 'stomach'. The metrics used here are best dice score, best jaccard value, train loss, valid dice, valid jaccard and valid loss. After implementing UNet++ model on each dataset, the outcome was pretty close to each other. The best result came when whole dataset was used. The separate dataset outcome was close, but they were not better than the whole dataset outcome. This is due to the fact that, in case of whole dataset, highest number of diverse data was used to train the model. In case of other three datasets, the number of data used to train the model was lower, hence degrading the outcome.

VI. Conclusion

In this article, we have been looking into the use of modern segmentation methods together with the GI tract image segmentation as a central topic. GI tract segmentation is a very important factor in right diagnosis and treatment of various gastrointestinal diseases like ulcerative colitis, Crohn's disease and gastrointestinal cancers.

Using our defined research, we have been able to establish that the U-Net and U-Net++ architectures have performed well in the job of segmenting the images. Such networks have demonstrated promising performance in correctly segmenting objects of interest in addition to enhancing the general understanding of health professionals and research workers.

The U-Net++ architecture, accompanied with the deep supervision mechanism that incorporates nested skip connections, has became a quite effective tool for image segmentation for GI tract. Through stacking features from various hierarchical levels and incorporating regularization into training, U-Net++ helps to improve segmentation accuracy and robustness, even when the anatomical form and imaging noise are present. In order to move forward, though, more research is required which takes into account the various problems and opportunities surrounding the segmentation of the GI tract image. We propose the development of multimodal imaging data integration algorithms, real-time segmentation algorithms for the endoscopic procedures, and transfer learning investigations for the pre-trained models and limited data. Our review also underscores the dataset diversity and model generalization as being crucial in attaining accurate segmentation outcomes. Through trained model on datasets acquired from various clinical centers and imaging modalities our segmentation approaches show good performance and generalization ability. In sum, the development in gastrointestinal tract image segmentation is a great opportunity in order to improve the diagnosis and treatment planning as well as to give a better quality of life to patients in gastroenterology. Through adoption of advanced separation approaches and the collaboration in multidisciplinary fields, we can form the path for the success of personalized medicine in the field of GI imaging as well as in general medicine.

Acknowledgment

This work was supported by Prof. Dr. Shaikh Anowarul Fattah, professor, department of EEE, BUET and Shahed Ahmed, lecturer, department of EEE, BUET. We also share our gratitude towards the author of the papers we've cited and towards the open source community which enabled us to get resources online. Last but not the least, we are grateful to our family and friends for their unconditional support.

References

[1] Hellier MD, Williams JG. The burden of gastrointestinal disease: implications for the provision of care in the UK. Gut. 2007 Feb;56(2):165-6.

[2]] Ramzan M, Raza M, Sharif MI, Kadry S. Gastrointestinal Tract Polyp Anomaly Segmentation on Colonoscopy Images Using Graft-U-Net. J Pers Med. 2022 Sep 6.

[3] D. Jha et al., "A Comprehensive Study on Colorectal Polyp Segmentation With ResUNet++, Conditional Random Field and TestTime Augmentation," in IEEE Journal of Biomedical and Health Informatics, June 2021,

[4] Ramzan M, Raza M, Sharif MI, Kadry S. Gastrointestinal Tract Polyp Anomaly Segmentation on Colonoscopy Images Using Graft-U-Net. J Pers Med. 2022 Sep 6.

[5] Kyeong-Beom Park , Jae Yeol Lee, SwinE-Net: hybrid deep learning approach to novel polyp segmentation using convolutional neural network and Swin Transformer Journal of Computational Design and Engineering, 2022.

[6] Y. Huang, D. Tan, Y. Zhang, X. Li and K. Hu, "TransMixer: A Hybrid Transformer and CNN Architecture for Polyp Segmentation," 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Las Vegas, NV, USA 2021.