IRES Report Summer 2024: Semantic Segmentation and Classification of OCT Colorectal Polyp Images

1st Gabrielle Miller
KIOS
Texas A and M University
College Station, USA
gabrielle._miller@tamu.edu

2nd Costas Pitris
KIOS
University of Cyprus
Aglantzia, Cyrpus
pitris.constantinos@ucy.ac.cy

3rd Andreas Spanias
SenSIP Center, ECEE
Arizona State University
Tempe, USA
spanias@asu.edu

***Abstract*—Colorectal cancer, being the second leading cause of cancer-related deaths, necessitates effective population screening methods, such as colonoscopy. Traditional histopathologic analysis of biopsies obtained during colonoscopy, although effective, is both costly and time-consuming, incurring a significant burden on the healthcare system. One approach to reduce this load is to "leave in situ" small benign masses and remove the ones with malignant potential. To achieve this goal, a highly accurate, in vivo, evaluation of the polyps is required. Here, a novel methodology is proposed, utilizing deep learning algorithms, to segment and classify polyps from en face ex vivo Optical coherence tomography (OCT) images. A DC-UNet based model is used to classify samples as either benign (normal and hyperplastic) or malignant potential (adenoma and sessile serrated adenoma) samples. The dataset consisted of 143 en face OCT images which were annotated by a histologist to create masks of the ground truth. Image segmentation was then performed using DC-UNet. The algorithm was also trained to classify the images. The training accuracy was 0.9808. Additionally, the model achieved a dice coefficient (F1 score) of 0.5923, Jaccard score of 0.4234, prediction accuracy of 0.72, sensitivity 0.72, and specificity 0.66. These preliminary results indicate that OCT has the potential to increase the accuracy of early detection of colon malignancies. In addition, with further improvements in the future, including a larger in vivo study, OCT can also become the enabler of a leave-in situ approach that can significantly reduce the burden of colonoscopies on the healthcare system.**

—

***Index Terms*—Optical Coherence Tomography, Deep Learning, UNet, Colorectal Polyps, Leaving-In-Situ**

I. Introduction

Colorectal Cancer (CRC) is one of the fourth most common cancer and is the second most common cause of cancer deaths in 2024 according to the World Health Organization. The American Cancer Society has found that young adults, below 55, are being diagnosed with CRC at a significantly higher rate than previously seen. Furthermore, the number of individuals being diagnosed with advanced stage has increased from 11% to 20% since 1995. Colonoscopic screening for CRC is highly recommended and was found to decrease the mortality of CRC between 18% to 57% in the past two decades. In higher-income countries, the survival rates of CRC are much higher, again reflecting the importance of colonoscopies. Although colonoscopy is an effective screening technique, it misses 20-30% of adenomas, suggesting room for improvement. Furthermore, most polyps identified during colonoscopy are benign, but the current approach is to remove and evaluate all of them. That introduces a significant burden to the healthcare system, hence the "leave-in-situ" approach, i.e. not resecting polyps less than 5 mm, is being explored. If a polyp is larger than 5 mm, it will still need to be removed. If it is classified as having malignant potential, then further screening and treatment would be advised. Optical Coherence Tomography (OCT) can be used to both improve the accuracy of colonoscopy but also enable "leave-in-situ" policies. Currently, OCT is commonly used in ophthalmology and cardiology, but its application to other areas is still lacking. Here, a novel methodology is proposed, utilizing deep learning algorithms, to segment and classify polyps from en face ex vivo Optical coherence tomography (OCT) images. A DC-UNet based model is used to classify samples as either benign (normal and hyperplastic) or malignant potential (adenoma and sessile serrated adenoma) samples.

II. Methodology

A set of 143 en face OCT images of colorectal polyps were collected ex vivo, immediately after colonoscopic resection. A Fourier domain OCT system was used, with a center wavelength of 1310 nm, an axial resolution of 14 m and an A-Scan rate of 100 kHz. The data were collected at the Massachusetts General Hospital. The en face images were created from volumes of 1000x1000x10000 pixels by automatically segmenting the images, removing the top surface, which contained multiple back-reflections, summing to a depth of 450 m, and finally normalizing. The images were annotated by an expert histopathologist as being normal, hyperplastic, adenoma, or sessile serrated adenoma (SSA). The classification was reduced to two-classes by grouping normal and hyperplastic as benign and adenoma and SSA as malignant potential.

U-Net is a widely used deep convolutional neural network designed specifically for medical image segmentation, addressing the challenges posed by small and irregularly shaped regions of interest in medical scans. The network begins with an encoder that captures the image context through a series of 3x3 convolutions followed by ReLU activations and 2x2 max pooling with a stride of 2 for down sampling. At

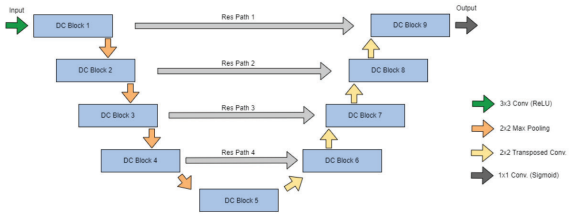

Fig. 1: DC-UNet Model U-Net Architecture which includes a dual channel efficient CNNs, in each Block, for medical image segmentation.

each step, the feature map size increases (64, 128, 256, 512, 1024), allowing the network to learn complex features. The bottleneck, transitioning from the encoder to the decoder, includes a 3x3 convolution followed by ReLU activation without down sampling6. The decoder then performs precise localization by up sampling the feature maps, followed by a 2x2 convolution that halves the number of feature channels. The concatenated feature maps are passed through two 3x3 convolutions with ReLU activations. The final output layer is a 1x1 convolution that reduces the output channels to the number of classes for pixel-wise classification. The U-Net's unique skip connections and symmetric architecture are crucial for accurate medical image segmentation, enhancing the network's ability to learn and retain detailed spatial information6. DC-UNet (Fig.1), is an enhanced model that incorporates a MultiResUNet architecture with inception blocks, which aid in extracting more features from the images. Additionally, a ResPath is added to the skip connections, and the dual-channel block replaces the residual connections in both the encoder and decoder. By incorporating these advanced features, the DC-UNet achieves better results than its predecessors, offering improved performance in detecting and classifying abnormal tissue6. The en face images and the annotation maps were fed to a DC-UNet model. The training was performed using the Adam optimizer for 10 to 50 epochs. 80 % of images were used as training and 20% for testing.

The classification was achieved using an encoder-decoder structure within the UNet. The encoder acts as a feature extractor, utilizing convolutional layers followed by ReLU activation and max-pooling layers to highlight specific sections of the scans that appear abnormal. These features are then passed to the decoder, where they are upsampled and concatenated with corresponding features from the encoder via skip connections to enhance accuracy. The decoder reconstructs the image, and at the final layer, a 1x1 convolution is used to map each pixel to a specific class. In this test, the model was configured to predict a single class, identifying only abnormal tissue. This approach allowed for pixel-by- pixel classification, ultimately creating a mask that highlights abnormal tissue.

III. Experimental Results

Detailed analysis of the results revealed that the model excelled at predicting normal tissue. This success indicated the model's capability to distinguish the benign tissue types which are characterized by areas of regularly arranged, uniform, villi. Dividing the segmentation into two distinct steps, i.e. separately for benign and malignant potential regions, the malignant potential regions were also effectively identified. The differences between the ground truth and the predicted regions were expected since the histopathologist annotated regions of "high confidence" only, to assure effective training of the algorithm. During classification, each epoch required approximately 100 to 150 seconds for training on a dual-core 3.3GHz computer. The training accuracy achieved was 98.08%, the prediction accuracy was 72%, the sensitivity was 72% and specificity was 66%. Additionally, the model achieved a dice coefficient (F1 score) of 0.5923 and a Jaccard score of 0.4234. The dice coefficient evaluates the overlap between the ground truth and predictions, emphasizing precision and recall. Conversely, the Jaccard score measures the intersection over the union of the predicted segmentation and the ground truth.

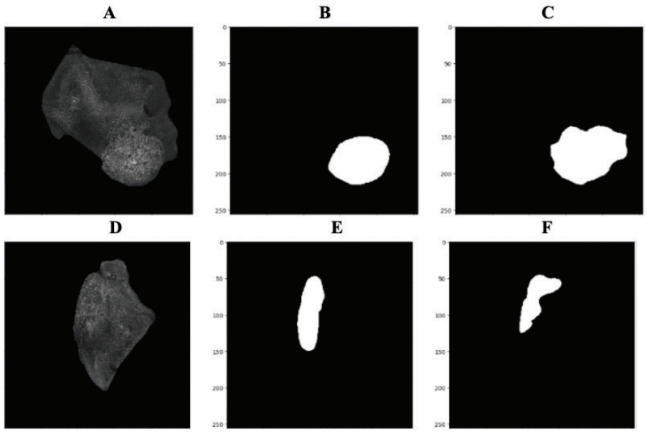

Fig. 2: Example results, (A) En face OCT image of a polyp with an adenomatous region. (B) The malignant potential area as marked by an expert histopathologist. (C) The prediction of the DC-UNet. (D)-(F) The same for a polyp with SSA. Each en face image image is 1000 x 1000 pixels of 5 x 5 mm.

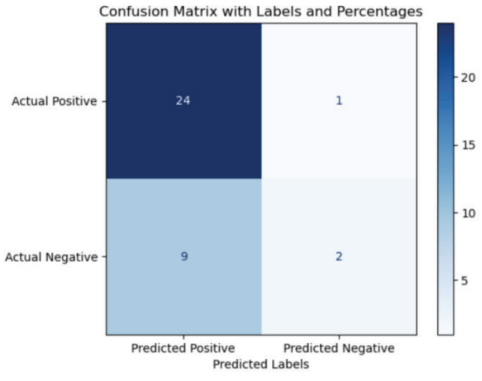

Fig. 3: Confusion matrix representation true positive, false positive, false negative, true negative.

IV. Conclusions

These initial results are promising, indicating significant potential for further modifications to the DC-UNet model. The demonstrated ability to recognize abnormal tissue is encouraging and, also, suggests that additional training could enhance performance. Later tests will include the addition of multi-spectral images and adding more channels to the DC-UNet which should increase accuracy. The initial runs were conducted on a standard PC; therefore, utilizing a deep-learning- optimized computer and further refining the model are expected to yield even better outcomes.

Acknowledgment

Support for this project was provided by NSF IRES Grant Award #1854273.

References

[1] "Seer Cancer Stat Facts." SEER, seer.cancer.gov/statfacts/. Accessed 3 July 2024.

[2] X. Llor, "Why Are Colorectal Cancer Rates Rising Among Younger Adults?" Yale School of Medicine. [Online]. Available: https://medicine.yale.edu/news-article/with-colorectal-cancer-rates-rising-among-younger-adults-a-yale-cancer-center-expert- explains-there-may-be-more-factors-behind-this-worrisome-trend/. Accessed: July 3, 2024.

[3] B. J. Elmunzer et al, "Comparing the effectiveness of competing tests for reducing colorectal cancer mortality: A network meta- analysis," Gastrointest. Endosc., vol 81, no. 3, 2015, doi: 10.1016/j.gie.2014.10.033.

[4] T. Pasricha, MD, and L. S. Friedman, MD, "How Well Do Colonoscopies Prevent Colorectal Cancer? What You Need to Know," Harvard Health, 18 Oct. 2022. [Online]. Available: www.health.harvard.edu/blog/how-well-do-colonoscopies-prevent-colorectal-cancer-what-you-need-to-know-202210182834.

[5] Fujimoto JG, Pitris C, Boppart SA, Brezinski ME. Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy. Neoplasia. 2000 Jan-Apr;2(1-2):9-25. doi: 10.1038/sj.neo.7900071. PMID: 10933065; PMCID: PMC1531864.

[6] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

[7] Ange Lou, Shuyue Guan, Murray Loew, "DC-UNet: rethinking the U-Net architecture with dual channel efficient CNN for medical image segmentation," Proc. SPIE 11596, Medical Imaging 2021: Image Processing, 115962T (15 February 2021); https://doi.org/10.1117/12.2582338