

ACCURATE COLON SEGMENTATION USING 2D CONVOLUTIONAL NEURAL NETWORKS WITH 3D CONTEXTUAL INFORMATION

Samir Harb^{1,2} A. Elsayed¹ M. Yousuf^{1,3} I. Alkabbany^{1,4} A. Ali¹ S. Elshazley⁵ A. Farag¹

¹ Computer Vision and Image Processing Laboratory (CVIP), University of Louisville, Louisville, KY

² Higher Technological Institute, 10th of Ramadan City, Egypt

³ Faculty of Engineering, Ain Shams University, Cairo, Egypt

⁴ Faculty of Engineering, Assiut University, Assiut, Egypt

⁵ Kentucky Imaging Technologies, Louisville, KY

ABSTRACT

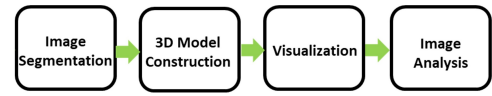
This study introduces an innovative framework designed specifically for accurate colon segmentation in abdomen CT scans, tackling the distinct challenges inherent to this task. Building upon well-established 2D segmentation models, our architecture adeptly incorporates 3D contextual information via a novel method that generates an attention map for a given slice by considering its neighboring slices in a sequence. Our approach accomplishes effective colon segmentation without requiring complex 3D convolutional neural networks (CNNs) or Long Short-Term Memory (LSTM) networks by combining 2D CNNs. Validated on a dataset of 98 CT scans from 49 patients, the architecture exhibits notable performance, successfully capturing nuanced details crucial for precise colon segmentation. The experiments encompass a thorough examination of model selection and cross-validation, providing valuable insights into the efficacy of our proposed approach. The outcomes underscore the potential for streamlined colon segmentation in medical imaging by judiciously integrating 2D and 3D information, employing solely 2D networks, and mitigating challenges associated with 3D networks. The code for model architecture is available at: <https://github.com/Samir-Farag/ICIP2024.git>

Index Terms— CRC, CTC, colon segmentation, deep learning.

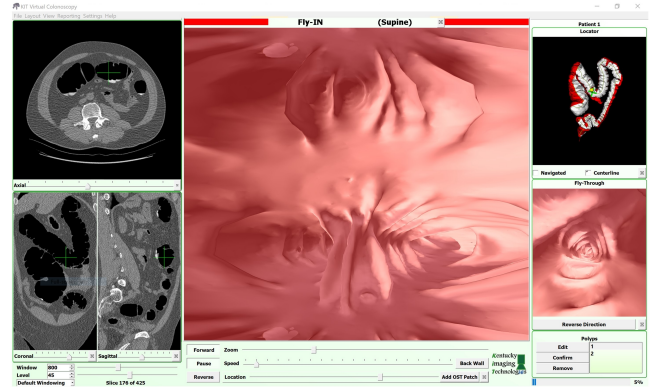
1. INTRODUCTION

Colorectal cancer (CRC) is a malignancy characterized by the uncontrollable proliferation of abnormal cells within the colon or rectum. CRC ranks as the third most frequently diagnosed cancer in the United States [1, 2]. To reduce mortality rates due to CRC, routine CRC screening is advised for people with average risk starting at the age of 45, see the American Cancer Society (ACS) guidelines for CRC [3]. CRC screening aims at identifying precancerous polyps

This work has been funded by NSF grant 2124316.



(a) Typical four-step CTC system, starting from CT to polyp detection



(b) GUI for Research Platform for CTC

Fig. 1. CTC Platform: (a) Typical pipeline; (b) CTC components: 2D CT slices, 3D model reconstruction and visualizations: Fly-In and Fly-Through.

before they progress into cancer and many CRC screening methods are available today, see [4] for a comprehensive list. These methods differ among them in terms of sensitivity, specificity, safety, invasivity, feasibility, and affordability with colonoscopy being the gold standard for CRC screening and adenoma identification.

Computed tomographic colonography (CTC) is a noninvasive screening method that does not necessitate sedation with good safety and ease of performance [5, 4]. CTC costs less compared to colonoscopy and its disadvantages diminish compared to the risk of operation and hospitalization for colonoscopy [6, 4]. Regarding the risk of radiation exposure, CTC has a significant benefit-to-risk ratio [7, 4].

Our team has an extensive experience (e.g., [8, 9, 10]) in developing CTC approaches. The CTC framework consists of four-steps, as shown in Fig. 1-(a), starting from an abdom-

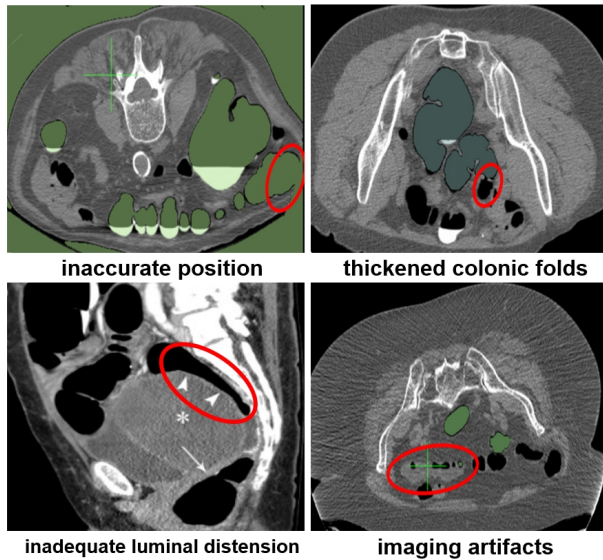


Fig. 2. Examples of segmentation challenges due to inaccurate position, thickened colonic folds, and other artifacts.

inal CT scan of a prepped patient. These steps are: 1) Image segmentation to isolate the lumen from the rest of tissues in the abdomen (e.g., the liver, lungs and small intestines, in addition to addressing uncertainties of CT acquisition); 2) 3D model building to construct the colon, which also involves extracting the centerline as a datum for visualization, and the registration of supine and prone CTC scans; 3) Visualization to present the lumen on radiology stations with details in 3D and corresponding 2D CT, Fig. 1-(b), in addition to functionalities for polyp editing; and 4) Analysis step which performs polyp detection, classification and archiving, and preparation of a full patient record.

The segmentation step is considered crucial in the aforementioned pipeline since errors in this step will propagate to the subsequent steps, moreover, the outcome of any subsequent step is indeed another representation of the information contained in the segmented image obtained in this step. This motivates us to focus on the segmentation step and enhance its performance. Enhancing the segmentation should entail performance gain in the subsequent steps.

Accurate segmentation of colons from abdomen CT scans is a challenging task, e.g., Fig. 2, due to: 1) colons are highly variable with asymmetric topology [11] and variable location for different CT images [12]; 2) the presence of Hounsfield intensity regions consisting of soft tissues, air regions which are similar to others gas-filled organs (e.g., small intestine), and high attenuation structures which are similar to bones; and 3) imperfect preparation of the patient, e.g., the presence of residual stool, lesions, which results in disconnected colon segments. These challenges may confuse segmentation algorithms and thus could hinder the accurate segmentation [12].

Colon segmentation approaches that have been reported

in the literature could be grouped into two main categories: (1) classic segmentation approaches, which typically employ techniques such as Markov Random Field model [8], edge detection, region growing and division [13, 14, 10]; and (2) deep learning (DL) approaches, see [15, 16, 17, 18] for more details. Compared to classical techniques, DL approaches achieve greater accuracy and often require less expert analysis and hand-crafted elements. DL approaches exploit the available data to learn complicated high-level characteristics that can be used for segmentation. On the other hand, classical approaches tend to focus on low-level traits that may not be as helpful for segmentation. Although DL approaches are successful segmentation tools, Jakob Wasserthal et al. [18], who developed the state-of-the-art (SOTA) segmentation approach, Totalsegmentator, reported that the colon posed the most significant challenges, with a failure rate of $\sim 35\%$ of cases. This failure primarily resulted from difficulties in accurately segmenting small parts of the colon.

In this work, we exploit the success of deep learning techniques in medical image segmentation and present a deep learning approach for automatically segmenting the colon from abdominal CT images. The primary contributions of our work can be succinctly outlined as follows:

- i Develop a novel 2D DL architecture that adeptly incorporates 3D contextual information, as shown in Fig. 3.
- ii Use a custom loss function to accurately train the colon segmentation model.
- iii Create a densely-annotated dataset of 49 patients (98 scans; prone and supine) for colon segmentation.

2. PROPOSED APPROACH

In the medical imaging literature, the decision between employing two-dimensional (2D) and three-dimensional (3D) deep learning models holds paramount importance. While 2D-based models adeptly handle individual CT images, offering computational efficiency and adaptability to irregular sampling, 3D models excel in capturing interslice context at the cost of heightened computational demands due to an expanded parameter space. It's noteworthy that applying 3D networks to process CT datasets may entail downsampling, potentially compromising crucial information.

2.1. A 2D meta-architecture

Considering the delicate balance between computational efficiency, model intricacy, and dataset requisites, we have chosen 2D CNNs as the preferred approach for colon segmentation from CT abdominal images. This decision seeks to strike a balance between computational practicality, segmentation accuracy, and the preservation of essential details in CT scans.

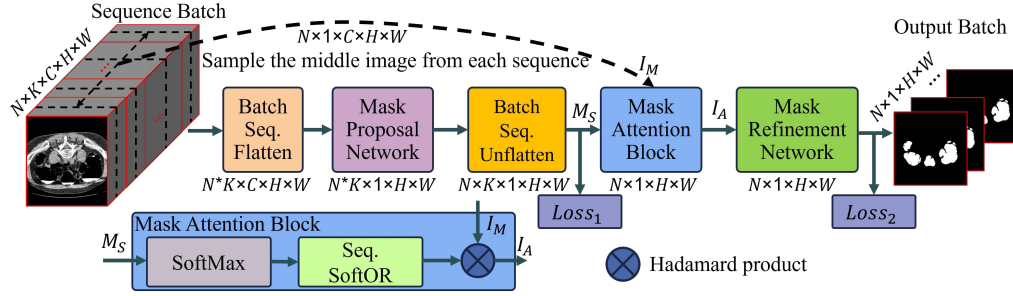


Fig. 3. Proposed Architecture: It consists of a Batch Sequence flatten block, a Mask Proposal Network, a Batch Sequence Unflatten Block, a Mask Attention Block, and a Mask Refinement Network. The Mask Attention Block consists of a softmax layer followed by a sequence softOR and a Hadamard product. Each of the Mask Proposal Network and the Mask Refinement Network can be any of-the-shelf segmentation model.

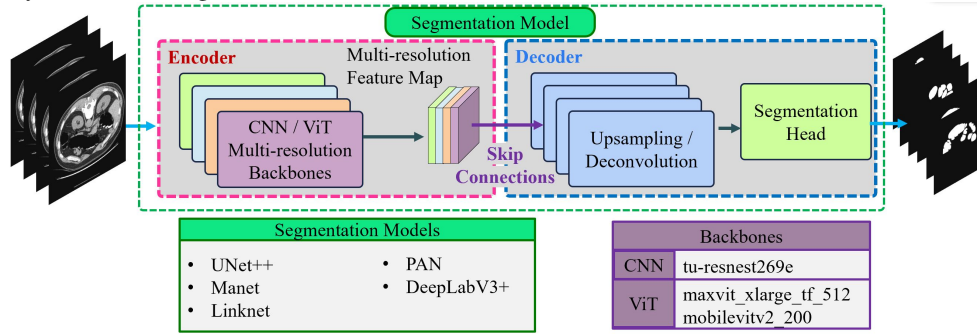


Fig. 4. A meta-architecture consists of encoder and decoder. The encoder has a CNN-based backbone which generates multi-resolution feature maps that are passed to the decoder via skip connections. The decoders fuses these features through upsampling/deconvolution blocks to generate a final feature map. The segmentation head uses this feature map to generate a predicted segmentation mask. During training, the custom dice uses the predicted and ground truth masks to update the network weights via back propagation.

Numerous segmentation architectures have been proposed, with the encoder-decoder architecture demonstrating significant success in various segmentation tasks. Fig. 4 illustrates an overview of the meta-architecture for a generic segmentation model, which consists of an encoder and a decoder. The encoder has a CNN backbone which generates a multi-resolution feature map from a given input image or image sequence. These feature maps are passed to the decoder via skip connections which use multiple upsampling blocks that fuse the multi-scale features and generate a final feature map. The final feature map is passed to the segmentation head to generate a segmentation mask that matches the input image resolution.

2.2. Proposed architecture

Most of the proposed segmentation architectures in the literature target 2D images in contrast to 3D image sequences. While the colon can be segmented effectively utilizing only information from 2D CT images, in this work, we introduce a straightforward yet effective novel architecture that harnesses

the power currently available 2D state-of-the-art segmentation model, while efficiently incorporating 3D information. The proposed approach is rooted in the idea that, rather than designing a completely new architecture, we should leverage established off-the-shelf building blocks and efficiently combine them to address our specific problem. As shown in experiments, this architecture can capture subtle issues that arise when using only 2D information to segment the colon.

Figure 3 depicts our proposed architecture for colon segmentation. It consists of main five blocks, a Batch-Sequence Flatten (BSF) block, a Mask Proposal Network (MPN), a Batch-Sequence Unflatten (BSU) block, a Mask Attention (MA) block, and a Mask Refinement Network (MRN), and two loss functions, with the Mask Attention block being the most important one.

Batch-Sequence Flatten (BSF) block: The proposed model takes a batch of CT image sequences I , as an input, with size $N \times K \times C \times H \times W$, where N, K, C, W, H are the batch size, the sequence length, the number of channels, the image width and the image height, respectively. The BSF block converts this batch to another batch with an equivalent

size of $N * K$ to prepare it for the MPN.

Mask Proposal Network (MPN): The MPN accepts the flattened batch and generates a batch of the corresponding proposed masks which then is converted back to a batch of mask sequences M_S by the Batch-Sequence Unflatten block.

Mask Attention (MA) block: The MA block receives the proposed mask sequences which are passed through a softmax layer for conversion from logits to probabilities. After that, each mask sequence is converted to only one attention mask corresponding to the middle image of the sequence. the conversion is done by summing all masks per sequence pixel-wise. This summation is similar to the logical OR operation for binary images, we call it sequence soft-OR. At the same time, the middle image from each sequence is sampled and passed to the MA block where it is attended by its corresponding attention mask using a Hadamard product to generate a batch of attended images I_A . The purpose of the mask attention block is to aggregate information from the neighborhood of each CT image by giving higher weight to pixels proposed by the MPN as being colon in this neighborhood. The motivation behind this is that the anatomy of the human body does not have abrupt changes within a small neighborhood around each CT slice and thus if a pixel in the current CT slice is colon it is most likely that this pixel will be colon within the next or the previous few CT slices. The word few here is reflected in the sequence length $K > 0$ which gives a neighborhood of $K/2$ around the current slice and is considered a hyper-parameter of the model with $K = 1$ is equivalent to segmentation using only 2D information.

Mask Refinement Network (MRN): Finally, the last stage of the pipeline MRN accepts the attended batch of CT images I_A and produces its final corresponding batch of segmentation masks. The MPN and the MRN can be any off-the-shelf 2D-segmentation model.

One important thing to note is that this architecture is totally 2D with a trick inside the mask attention block that aggregates information from the neighborhood of the current mask to generate a single attention mask for the current CT slice. This approach is efficient in terms of model size and computational cost compared to 3D CNNs and simple to train compared to LSTM.

Finally, two separate loss functions are used, one to force the MPN to propose accurate masks and the other to force the MRN to generate accurate segmentation masks from the attended images I_A . The first loss compares the ground truth segmentation masks with the proposed masks while the second loss compares the ground truth masks with the final predicted masks.

2.3. Custom loss function

While most segmentation problems treat the objects to be segmented as a whole, i.e. all the pixels belonging to a certain class or object are treated as equally important, this should not

be the case for colon segmentation. The reason is that pixels at the boundary of the colon are more important than pixels at its interior since almost all colon's interior is filled with air which has a well-known Hounsfield value range, making it easy to segment by simple thresholding. Thus, to force the model to learn the intricate details of the colon boundary, we designed a per-pixel weighted dice loss that gives more weight to pixels near the boundaries of the colon (boundary weight). For each CT image pixel, a weight is assigned based on a weight map computed from each ground truth segmentation mask corresponding to this slice, see Fig. 5-(a, b). The purpose of this modification is to give more importance to pixels near the boundary of the colon region in the loss calculation. This is often necessary in medical imaging tasks where precise boundary detection is crucial. The loss function is defined by

$$\mathcal{L}(x, y, b) = \alpha \cdot DiL(x, y) + (1 - \alpha) \cdot BL(x, b), \quad (1)$$

where, x represents the predicted segmentation mask, y denotes the ground truth binary mask, b stands for the computed boundary map, α is the boundary weight, a hyperparameter controlling the trade-off between the dice loss, (DiL), and the boundary loss, (BL), enabling the prioritization of boundary pixels in the loss function to improve the precision of colon boundary detection. The DiL is defined by

$$DiL = 1 - \frac{2 \sum_{i=1}^N p_i g_i}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N g_i^2}, \quad (2)$$

where $p_i \in [0, 1]$ signifies the probability for the i -th pixel to be colon, and $g_i \in \{0, 1\}$ represents the ground truth for the same pixel, and N represents the total number of pixels [19]. Further, $BL(x, b) = \frac{1}{N} \sum_{i=1}^N b_i \cdot x_i$, where, N represents the total number of pixels, b_i and x_i are the values of the boundary map and predicted segmentation mask at pixel i , respectively. The boundary map b is calculated by finding the minimum Euclidean distance from each pixel (i, j) to any pixel on the boundary of the binary mask M . The distance transform assigns a higher value to pixels closer to the boundary, and a lower value to pixels in the interior of the binary mask. Where $b_{i,j}$ is the value in the boundary map at position i, j , that is the value in the distance map at the same position:

$$DM(M)_{i,j} = \min_{(x,y) \in \text{Boundary}(M)} \sqrt{(i-x)^2 + (j-y)^2}. \quad (3)$$

3. EXPERIMENTS

We conduct a series of experiments to rigorously evaluate the effectiveness of our proposed colon segmentation approach. Two pivotal questions are addressed: 1) How can we assess the performance of our model given that there is no standard benchmark for colon segmentation? 2) How should we

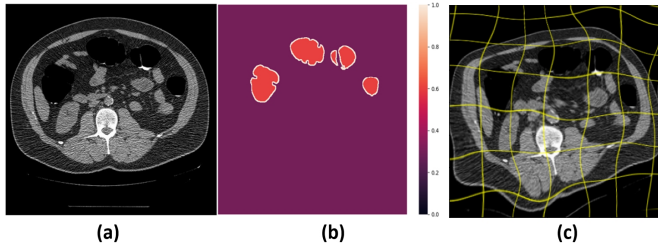


Fig. 5. An example of a CT slice (a) and its pixel-wise weight map (b) that is generated from the ground truth to force the network to learn the colon surface. An example of data augmentation with elastic deformation (c) for the same CT slice.

choose the MPN and the MRN? Unfortunately, most of the previous work, e.g., [12, 17] evaluated their models on either a private dataset or a public dataset whose annotation is still private. Given that in mind and to address the first question, we will pick a base model against which we will assess our model performance on our private dataset. Regarding the second question, we should choose the MPN and MRN in a way that allows a fair comparison with the base model. The goal of the first experiment is to investigate which SOTA segmentation architecture and backbone are optimum for the base model. The second experiment objective is to show that our proposed model can overcome the subtle issues encountered by the base model and to make sure that the model generalizes well over the whole dataset. Also, it's important to highlight that we conducted a comparative analysis by contrasting our proposed method with the nnUNet [20] which is the SOTA 3D-based method for medical image segmentation.

Dataset: We conducted our experiments on a private dataset provided by CTC experts from the University of Wisconsin. This dataset comprises scans of 49 patients in both supine and prone positions. Experts annotated the colon segments in these 98 CT scans. We normalized CT images with $\mu = 841.78$ and $\sigma = 1508.98$ before being fed into the model for training or validation, where μ and σ are the mean and standard deviation, respectively, of the raw Hounsfield values over the whole dataset, which is available upon request.

Base Model: Among the encoder-decoder architectures proposed in the literature, see [21] for a detailed review, UNet [22], UNet++ [23], Linknet [24], FPN [25], PSPNet [26], PAN [27], DeepLabV3 [28], and DeepLabV3+ [29] are considered the most successful ones. The brute force experimentation with all of these models to pick the most accurate base model is not feasible. However, UNet++ and DeepLabV3+ are updated versions of UNet and DeepLabV3, respectively, and thus perform better. PSPNet and FPN perform poorly compared to all other models, especially PSPNet performs badly for detecting small objects and producing accurate pixel-level masks. So, we will not experiment with these specific two models. Taking this into account, we are only left with four models, i.e., UNet++, Linknet,

PAN, DeepLabV3+. There are hundreds of available backbones that can be used with these selected four segmentation models and experimenting with all these variants is almost impossible. To tackle this problem, we will take the number of encoder parameters as the criteria for choosing backbones. This criterion is valid because as the number of parameters increases the model capacity increases and this gives the model the ability to learn complex tasks such as colon segmentation. Conducting a series of experiments with different architectures and backbones, Table 2, shows that UNet++ with resnet269e backbone gives the best performance for the base model. Now for fair comparison, we utilized UNet++ with resnet269e backbone for the MPN and UNet++ with resnet26d backbone for the MRN. The reason we picked a smaller backbone for the MRN is that its task is much easier since the heavy work has already been done by the MPN.

Evaluation: In the second experiment, we conducted a patient-independent 7-fold cross-validation for the base model with resnet269e encoder and UNet++ decoder as well as our proposed model. To enrich the training dataset, we incorporate deformable augmentations e.g., Fig. 5-(c), which is an appropriate augmentation in case of soft organs like colon. The total size of the dataset is 42,609 CT images; divided into 42 patients/ 7 patients on average 36,469/6140 images respectively for training/validation in every fold. Table 1 indicates the validation metrics for the base model compared to the proposed model. We employed different well-known metrics [30] such as the dice coefficient, Jaccard Coefficient, and Hausdorff Distance. These metrics quantify the pixel-wise agreement between the predicted and ground truth segmentation [31]. Additionally, we incorporated the average inference time per scan for each model as a trade-off consideration.

Due to computational burden, the 3D network was evaluated using low-resolution images. However, the 2D-based proposed model performs as accurately as the 3D network (Fig.6-e) without the need to downscale the images. Despite this, the base model encountered some subtle limitations, as depicted in Fig.6-c. The C-LSTM model exhibited the worst performance (shown in Fig.6-d), almost missing most details. In contrast, the proposed model managed to address these limitations effectively, as illustrated in Fig.6-f.

The primary challenge in traditional encoder-decoder networks lies in their inability to incorporate temporal information in a sequence of images such as colon CT scan. Therefore, we explore the fusion of C-LSTM with UNet by replacing the convolutional layers in the encoder section with C-LSTM layers, as described in [32]. However, its performance is still lower than the proposed approach, as shown in Table 1, because it has a larger number of parameters that should be optimized and this hinders the network learning especially for long and high-resolution image sequences.

Technical and Implementation details: To train all models, an AdamW optimizer [33] is used with an initial

Table 1. Evaluation of volume segmentations for nnUNet, Base Model (UNet++/resnest269e), C-LSTM, and the proposed model using the metrics reported in [30]. The dataset size for train/val: 29,430/7,786. INFTIME stands for average inference time per scan.

Model	DICE	JACRD	bAVD	KAPPA	ICCORR	MUTINF	SNSVTY	SPCFTY	INFTIME
nnUNet [20]*	98.79%	97.61%	0.0133	98.73%	98.79%	25.78%	98.85%	99.94%	7 min
Base Model	97.44%	95.75%	0.3136	97.39%	97.47%	22.72%	98.01%	99.89%	6 min
C-LSTM	89.16%	80.68%	0.39	88.62%	91.10%	21.30%	96.38%	99.08%	10 min
Ours	98.76%	97.60%	0.03	98.70%	98.76%	22.90%	98.60%	99.90%	5 min

* Due to computation burden, it was evaluated on half the resolution.

Table 2. Validation Dice and JACRD Scores for Multiple Architectures with Different Backbones.

Architecture/Backbone	Val. Dice Score	JACRD
PAN/resnest269e	97.1%	95.46%
DeepLabV3+/resnest269e	97.35%	95.74%
Linknet/resnest269e	97.36%	95.96%
UNet++/resnest269e	97.44%	95.75%

learning rate set to 10^{-4} , and a weight decay of 3^{-2} . Moreover, A cosine annealing learning rate [34] is used with T_{max} is set to the number of batches per epoch. All models are trained with batch size equal to 8, and image size 512×512 except the C-LSTM-based network is trained with image size 128×128 . Moreover, we start training with a pre-trained model from PyTorch Segmentation Models Library [35]. All the training was performed on Nvidia TITAN RTX 24 Gb.

Discussion: From the conducted experiments, we can conclude that, with a suitable loss function and appropriate data augmentation, colon segmentation can be done in 2D quiet efficiently using the existing off-the-shelf segmentation models and backbones, specifically using UNet++ with resnest269e as a backbone. However, the best performing model did miss some tiny details in the colon in some CT images (see Fig. 6), which could be reflected in colon 3D reconstruction. We also noted that the proposed model did succeed to get the missed parts in some of the neighboring slice. This suggests that incorporating 3D information via efficient 3D convolution e.g., [36] may address this limitation. One important rule to note is that splitting data between training and validation should be done per patient, i.e. no CT images from the same patient contribute to the training and validation dataset at the same experiment. This is critical since each two successive CT images of the same patient almost have the same information, and accordingly, when we violated this rule, we got a validation dice score of **99.4%**.

4. CONCLUSIONS

In this study, we introduced a novel architecture for colon segmentation in 3D medical imaging, seamlessly blending the capabilities of established 2D state-of-the-art segmentation models with efficient 3D integration. Extensive experiments

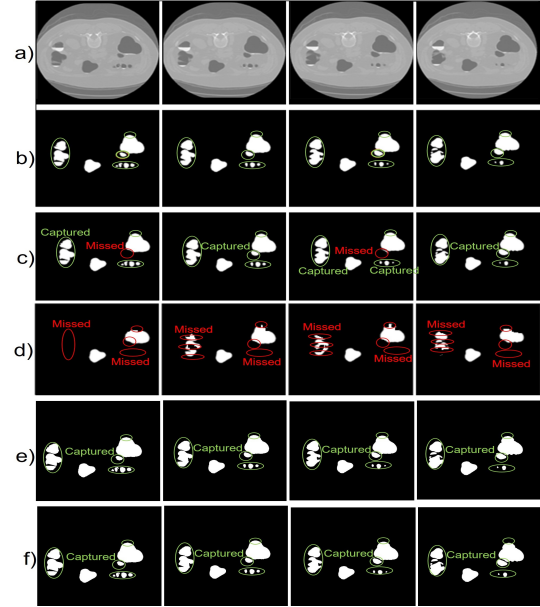


Fig. 6. Example for models limitations. (a) Original CT images, (b) Ground truth labels, and predicted labels for Base Model (c), C-LSTM (d), nnUNet (e), as well as Ours (f). Green circles indicate challenging regions that may be missed by a model. Both 3D-based approach nnUNet (e) and our 2D-based approach (f) successfully capture these regions.

on a manually annotated private dataset of 98 scans, encompassing both prone and supine positions, highlight the high performance of the proposed approach. We built on the UNet++ model with the resnest269e backbone, which outperforms other encoder-decoder architectures. Moreover, our proposed model exhibits a unique proficiency in restoring intricate details of the colon, challenging to be captured by the base model. This precision advancement holds great promise for downstream applications, particularly in 3D reconstruction. In contrast, alternative techniques like Convolutional Long Short-Term Memory (C-LSTM) proved demanding in training time and failed to deliver commensurate improvements. The incorporation of a custom loss function, emphasizing critical features like colon boundaries, further elevates our model's performance. Overall, our proposed method shows promising results, providing a robust foundation for future applications in clinical settings.

5. REFERENCES

- [1] A. Karaman, I. Pacal, A. Basturk, B. Akay, U. Nalbantoglu, S. Coskun, O. Sahin, and D. Karaboga, "Robust real-time polyp detection system design based on yolo algorithms by optimizing activation functions and hyper-parameters with artificial bee colony (abc)," *Expert Systems with Applications*, vol. 221, pp. 119741, 2023.
- [2] R. L. Siegel, N. S. Wagle, A. Cercek, R. A. Smith, and A. Jemal, "Colorectal cancer statistics, 2023," *CA: a cancer journal for clinicians*, vol. 73, no. 3, pp. 233–254, 2023.
- [3] "The american cancer society," <https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/acs-recommendations.html>, Accessed: 2021-12-28.
- [4] M. Jayasinghe, O. Prathiraja, D. Caldera, R. Jena, J. A. Coffie-Pierre, M. S. Silva, O. S. Siddiqui, and J. A. Coffie-Pierre Jr, "Colon cancer screening methods: 2023 update," *Cureus*, vol. 15, no. 4, 2023.
- [5] J. Sha, J. Chen, X. Lv, S. Liu, R. Chen, and Z. Zhang, "Computed tomography colonography versus colonoscopy for detection of colorectal cancer: a diagnostic performance study," *BMC Medical Imaging*, 2020.
- [6] S. Halligan, E. Dadswell, K. Wooldrage, J. Wardle, C. von W., R. Lilford, G. L. Yao, S. Zhu, and W. Atkin, "Computed tomographic colonography compared with colonoscopy or barium enema for diagnosis of colorectal cancer in older symptomatic patients: two multi-centre randomised trials with economic evaluation (the siggar trials).," *Health technology assessment (Winchester, England)*, vol. 19, no. 54, pp. 1–134, 2015.
- [7] D. J. Brenner and M. A. Georgsson, "Mass screening with ct colonography: should the radiation exposure be of concern?," *Gastroenterology*, vol. 129, no. 1, pp. 328–337, 2005.
- [8] M. Yousuf, I. Alkabbany, A. Ali, S. Elshazley, A. Seow, G. Dryden, and A. Farag, "An automatic colorectal polyps detection approach for ct colonography," in *IEEE ICIP*, 2023, pp. 2790–2794.
- [9] M. Mohamad, A. Farag, A. M. Ali, S. Elshazly, A. A. Farag, and M. Ghanoum, "Enhancing virtual colonoscopy with a new visualization measure," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 294–297.
- [10] D. Chen, R. Fahmi, A. A. Farag, R. L. Falk, and G. W. Dryden, "Accurate and fast 3d colon segmentation in ct colonography," in *2009 IEEE ISBI: From Nano to Macro*, pp. 490–493.
- [11] Z. Ravindran and N. S. Das, "Automatic segmentation of colon using multilevel morphology and thresholding," in *2021 IEEE Int. Conf. on Computer Communication and Informatics*, pp. 1–4.
- [12] K. Gayathri and R. Radhakrishnan, "Automatic segmentation of colon in 3d ct images and removal of opacified fluid using cascade feed forward neural network," *Comput. and Math. Methods in Med.*, 2015.
- [13] J. J. Nappi, A. H. Dachman, P. MacEaney, and H. Yoshida, "Effect of knowledge-guided colon segmentation in automated detection of polyps in ct colonography," in *Medical Imaging 2002: Physiology and Function from Multidimensional Images*. SPIE, vol. 4683.
- [14] D. K. Gayathri, R. Radhakrishnan, and K. Rajamani, "Segmentation of colon and removal of opacified fluid for virtual colonoscopy," *Pattern Analysis and Applications*, vol. 21, no. 1, pp. 205–219, 2018.
- [15] Gupta S. Koundal D. Zaguia A. Malhotra, P. and W. Enbeyle, "Deep neural networks for medical image segmentation," *Journal of Healthcare Engineering*, vol. 2022, 2022.
- [16] L. Guachi, R. Guachi, F. Bini, and F. Marinozzi, "Automatic colorectal segmentation with convolutional neural network," *Computer-Aided Design and Applications*, vol. 16, no. 5, pp. 836–845, 2019.
- [17] A. Akilandeswari, D. Sungeetha, C. Joseph, K. Thaiyalnayaki, K. Baskaran, R. Jothi R., H. Al-Lohedan, D. M. Al-Dhayan, M. Karnan, and K. Meansbo H., "Automatic detection and segmentation of colorectal cancer with deep residual convolutional neural network," *Evidence-Based Complementary and Alternative Medicine*, vol. 2022, 2022.
- [18] J. Wasserthal, H. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, M. Bach, and M. Segeroth, "Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images," *Radiology: AI*, vol. 5, no. 5, 2023.
- [19] R. Zhao, B. Qian, X. Zhang, Y. Li, R. Wei, Y. Liu, and Y. Pan, "Rethinking dice loss for medical image segmentation," in *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 851–860.
- [20] Fabian Isensee, Paul F. Jaeger, Simon A. Kohl, Jens Petersen, and Klaus H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [21] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3523–3542, 2021.
- [22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18. Springer, pp. 234–241.
- [23] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 1856–1867, 2019.
- [24] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE visual communications and image processing (VCIP)*, pp. 1–4.
- [25] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *IEEE CVPR*, 2017.
- [26] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *IEEE CVPR*, 2017.
- [27] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," *arXiv preprint arXiv:1805.10180*, 2018.
- [28] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.
- [29] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV*, 2018.
- [30] Abdel Aziz Taha and Allan Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC medical imaging*, vol. 15, no. 1, pp. 1–28, 2015.
- [31] R. R. Shamir, Y. Duchin, J. Kim, G. Sapiro, and N. Harel, "Continuous dice coefficient: a method for evaluating probabilistic segmentations," *arXiv preprint arXiv:1906.11031*, 2019.
- [32] A. Arbel and T. R. Raviv, "Microscopy cell segmentation via convolutional lstm networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1008–1012.
- [33] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
- [34] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.
- [35] Pavel Iakubovskii, "Segmentation models pytorch," https://github.com/qubvel/segmentation_models.pytorch, 2019.
- [36] Qianfei Zhao, Lanfeng Zhong, Jianghong Xiao, Jingbo Zhang, Yinan Chen, Wenjun Liao, Shaoting Zhang, and Guotai Wang, "Efficient multi-organ segmentation from 3d abdominal ct images with lightweight network and knowledge distillation," *IEEE Transactions on Medical Imaging*, vol. 42, no. 9, pp. 2513–2523, 2023.