47

Deep Learning-Based Colon Segmentation for Accurate Colorectal Polyps Detection

Samir Harb, Mohamed Yousuf, Ahmed Elsayed, Asem Ali, Salwa Elshazly, and Aly Farag

47.1 Introduction

Colorectal cancer (CRC) is a type of cancer characterized by an uncontrolled growth of abnormal cells in the colon or rectum, ranking as the third most common cancer in the United States [1]. The American Cancer Society (ACS) [2] recommends that screening for CRC should start at age 45. Early diagnosis and treatment of CRC has been shown to be most effective of treating this cancer with significant recovery rate (e.g., [3]).

There are various methods available for CRC screening, differing in sensitivity, specificity, safety, invasiveness, feasibility, and cost [4]. Optical colonoscopy is considered the gold standard method for CRC screening. On the other hand, CT colonography (CTC) is noninvasive and performed without sedation in less time and is quite safe (e.g., [4, 5]). It is also less expensive than colonoscopy, with its drawbacks being less significant compared to the risks associated with colonoscopy-related procedures and hospitalization [4, 6]. Despite concerns about radiation exposure, CTC offers a favorable benefit-to-risk ratio [4, 7].

The CTC pipeline is comprised of four distinct steps, as shown in Figure 47.1. First, image segmentation isolates the lumen from surrounding abdominal tissues. The second step involves the 3D construction of the colon. This enables extracting the centerline for visualization purposes and ensuring proper alignment of supine and prone CTC scans. The third step focuses on visualization, presenting the lumen on radiologists' stations with detailed 3D and corresponding 2D CT views. Finally, in the fourth and crucial step, an in-depth analysis is conducted. This involves polyp detection, classification, archiving of relevant data, and the meticulous preparation of a comprehensive patient record. Through these steps, the CTC framework enables screening for CRC and assists other functionalities such as training and image-guided interventions related to combating CRC. There exists a large literature on CTC (e.g., [8–10]).

Segmenting colons accurately from abdomen CT scans poses significant challenges, as depicted in Figure 47.2. First, colons exhibit highly variable and asymmetric topology [11], and their positions vary across different CT images [12]. Second, distinguishing colons from surrounding structures is complicated by the presence of Hounsfield intensity regions containing soft tissues, air regions resembling gas-filled organs like the small intestine, and high attenuation structures akin to bones. Lastly, patient preparation imperfections, such as residual stool and lesions, can lead to disjointed colon segments. These complexities may confound segmentation algorithms, impeding accurate results [12].

Biomedical image segmentation approaches in the literature could be grouped into two main categories: (1) classic segmentation approaches, which typically employ techniques such as Markov random field (MRF) models [8], edge detection, and region growing [10, 13, 14]; and (2) deep learning (DL) approaches; which casts the segmentation as a classification problem solved by convolutional neural networks (CNNs) identified by deep learning techniques (e.g., [15–18]). MRF-based methods exploit the connectivity of neighboring pixels/voxels that belong to the object of concern. DL approaches exploit the available data to learn complicated high-level characteristics that can be used for segmentation. Classical approaches tend to focus on low-level traits that may not be as helpful for segmentation. Although DL approaches have shown better performance in segmentation, in general, Wasserthal and colleagues [18] reported that the colon posed the most significant challenges, with a failure rate of ~35% of cases. This failure primarily resulted from difficulties in accurately segmenting small parts of the colon.

Most DL segmentation approaches depend on a meta-architecture, which consists of encoder and decoder. The encoder has a CNN-based backbone that generates multi-resolution feature maps that are passed to the decoder via skip connections. The decoders fuse these features through deconvolution blocks to generate a final feature map. The segmentation head uses this feature map to generate a predicted segmentation mask. Numerous segmentation approaches, for example UNet [19], UNet++ [20], Linknet [21], FPN [22], PSPNet [23], PAN [24], DeepLabV3 [25], and DeepLabV3+ [26], have been proposed with the encoder-decoder architecture demonstrating significant success in various segmentation tasks.

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

DOI: 10.1201/9781003449942-53

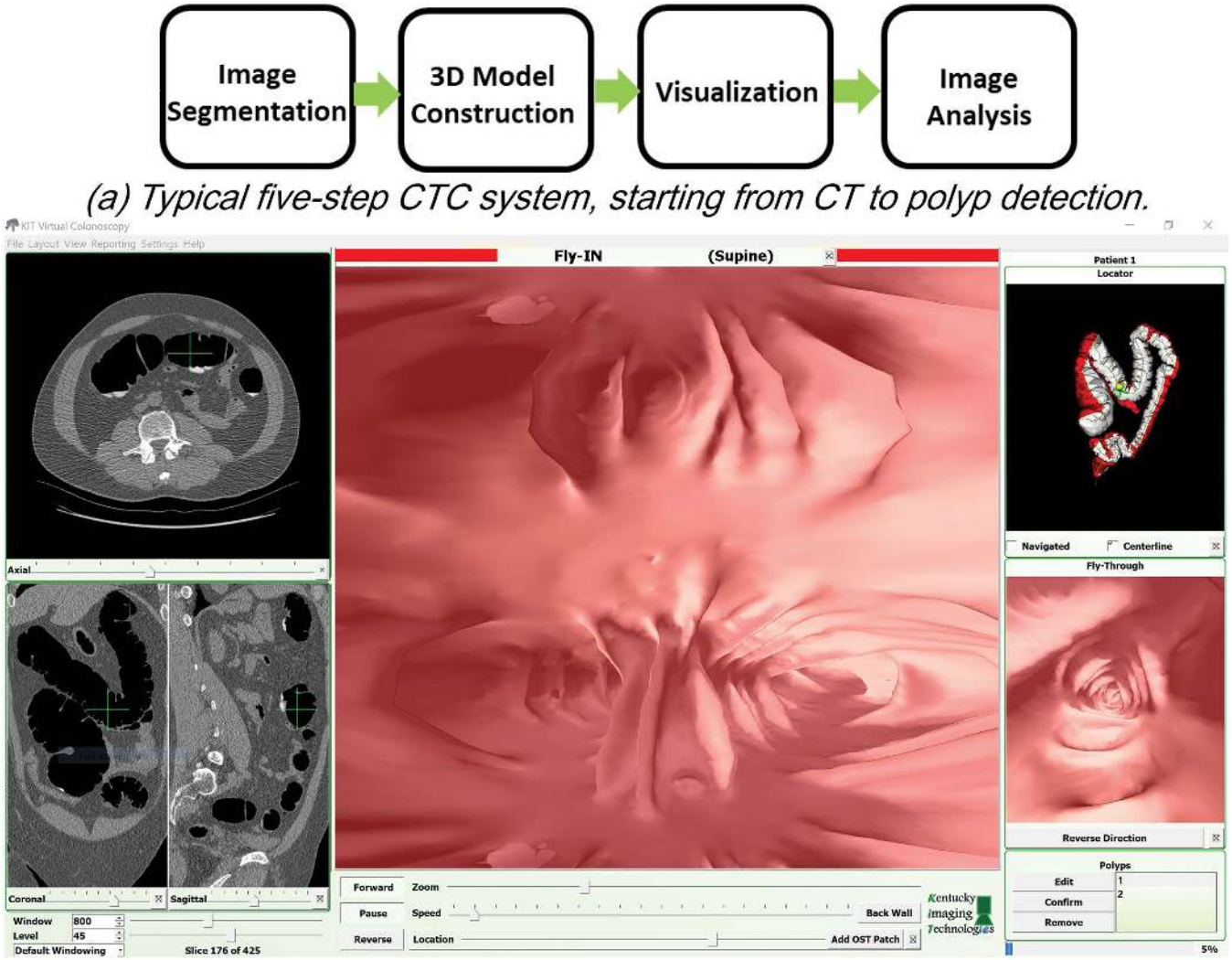

FIGURE 47.1 CTC platform: (a) typical pipeline; (b) CTC components: 2D CT slices, the 3D model reconstruction, and visualizations: Fly-In and Fly-Through.

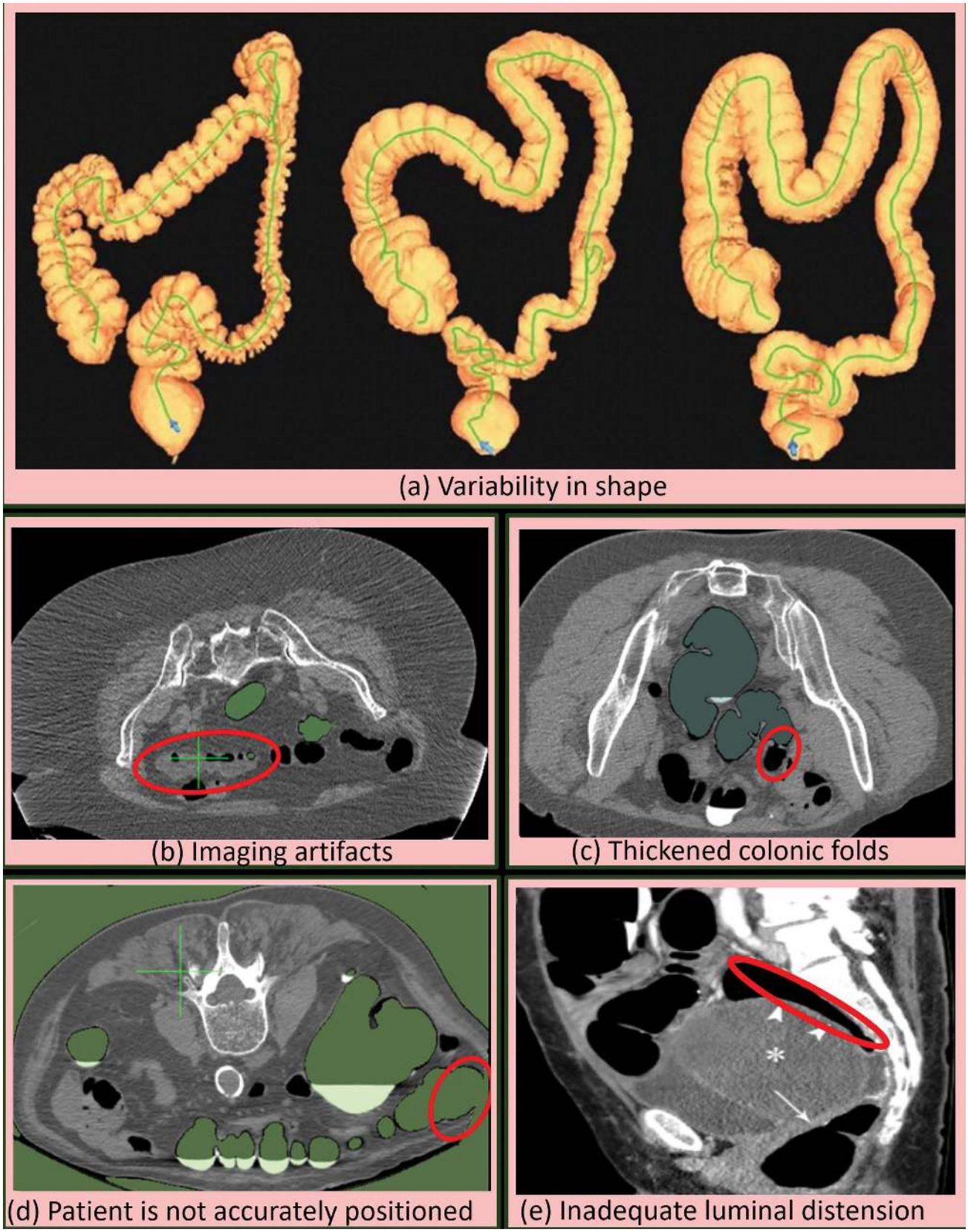

FIGURE 47.2 Examples of segmentation challenges due to inaccurate position, thickened colonic folds, and other artifacts.

cost of heightened computational demands due to an expanded parameter space. So, applying 3D networks to process CT datasets may entail downsampling, potentially compromising crucial information.

47.2 Materials

47.2.1 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 $=1508.98$ 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. This research study was conducted retrospectively using human subject data made available in open access by ACRIN, Walter Reed Medical Center, and the University of Wisconsin [27]. The research protocol is governed by the University of Louisville IRB No. 07.0252.

47.2.2 Hardware

The used server hardware for the training and validation has the following specs: Processor: AMD EPYC 7702 P 64-Core Processor, RAM size:264 GB, used GPU: two Nvidia TITAN X with 24 GB of VRAM each.

For the CT device to generate the dataset, the manufacturer was GE Medical Systems, with Kilo Voltage Peak = 120, slice thickness 1.25, and X-ray tube current 180 mA.

47.2.3 Software and Packages

The Ubuntu 20.04.6 LTS is the main operating system that had been used, alongside with Python 3.9 as the main programming language. Further, Numpy, pandas, PyDicom, and OpenCV libraries have been used for preparing the CT images before processing.

For the CNN-based segmentation approach, to train all models, an AdamW optimizer [28] is used with an initial learning rate set to 10^{-4}, and a weight decay of 3^{-2}. Moreover, a cosine annealing learning rate [29] is used with T_{max} is set to the number of batches per epoch. All models are trained for ten epochs with batch size equal to eight, and image size 512×512. Moreover, we start training with a pretrained model from PyTorch Segmentation Models Library [30].

47.3 Methods

In this chapter, we present two quantitative algorithms for colon segmentation in CTC using MRF and DL.

47.3.1 MRF-Based Segmentation Approach

In a typical CTC scan, the lumen has two components: opacified fluid whose Hounsfield intensity is greater than 300 HU, and air, for which the characteristic peaks are almost at 1000 HU (e.g., [13]).

The problem is formulated as *maximum a posteriori* (MAP) estimation of a MRF, which involves finding the labeling that minimizes the following energy function $E(\mathrm{x})$ that combines both the spatial smoothness and data consistency to achieve an accurate segmentation of the colon in CT scans:

$$E(\mathbf{x}) = \sum_{\{i,j\}\in\mathcal{N}} V\left(x_i, x_j\right) + \sum_{i\in\mathcal{P}} D\left(x_i\right). \tag{47.1}$$

where $\mathcal{N}$ represents the set of neighboring pixel pairs (i, j), $V(x_i \mid x_j)$ is the potential function that penalizes label inconsistencies between neighboring pixels, and $D\left(x_i\right)$ is the data penalty term that measures how well the labeling x_i matches the observed image I_i. Minimizing energy function in Equation 47.1 using graph cut generates the optimal segmentation result.

First, we use expectation-maximization (EM) to calculate the empirical distributions of Hounsfield intensities in a DICOM volume $\mathbf{V}$, as shown in Figure 47.3b. This step is mandatory to identify the different classes, for example air, fat, fluid, and so on. Then, we identify colon regions using two thresholds, which are calculated using four empirical distributions of air, fat, muscle, and fluid. The threshold between air and fat and the other between muscle and fluid are used to generate O_{EM}. However, the initial segmentation results O_{EM} include noncolonic tissues, for example bones as shown in Figure 47.3c. To exclude these noncolonic components, we apply the region-growing (RG) algorithm, starting from the rectum, which can be easily identified as shown in Figure 47.3d. The RG algorithm is designed to generate an eroded colon O_{RG}. This guarantees that other organs, for example small intestine, are not merged with O_{RG}. Finally, to extract optimal segmentation, we exploit the pixel information in addition to its neighborhood to refine the initial segmentation O_{RG} through an optimization technique.

47.3.2 CNN-Based Segmentation Approach

Figure 47.4 illustrates a meta-architecture for a CNN-based classifier used for colon segmentation, consisting of an encoder and a decoder. The encoder has a CNN backbone that generates a multi-resolution feature map from a given input image or an 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.

CNN-based segmentation deals with two main issues: the architecture and the encoder. We use architectures that are versatile with different backbones, and use the rule of thumb that a backbone with more parameters equates to a higher model capacity and, consequently, can handle more complex tasks. These two constraints effectively limit the number of architectures/backbones we should test.

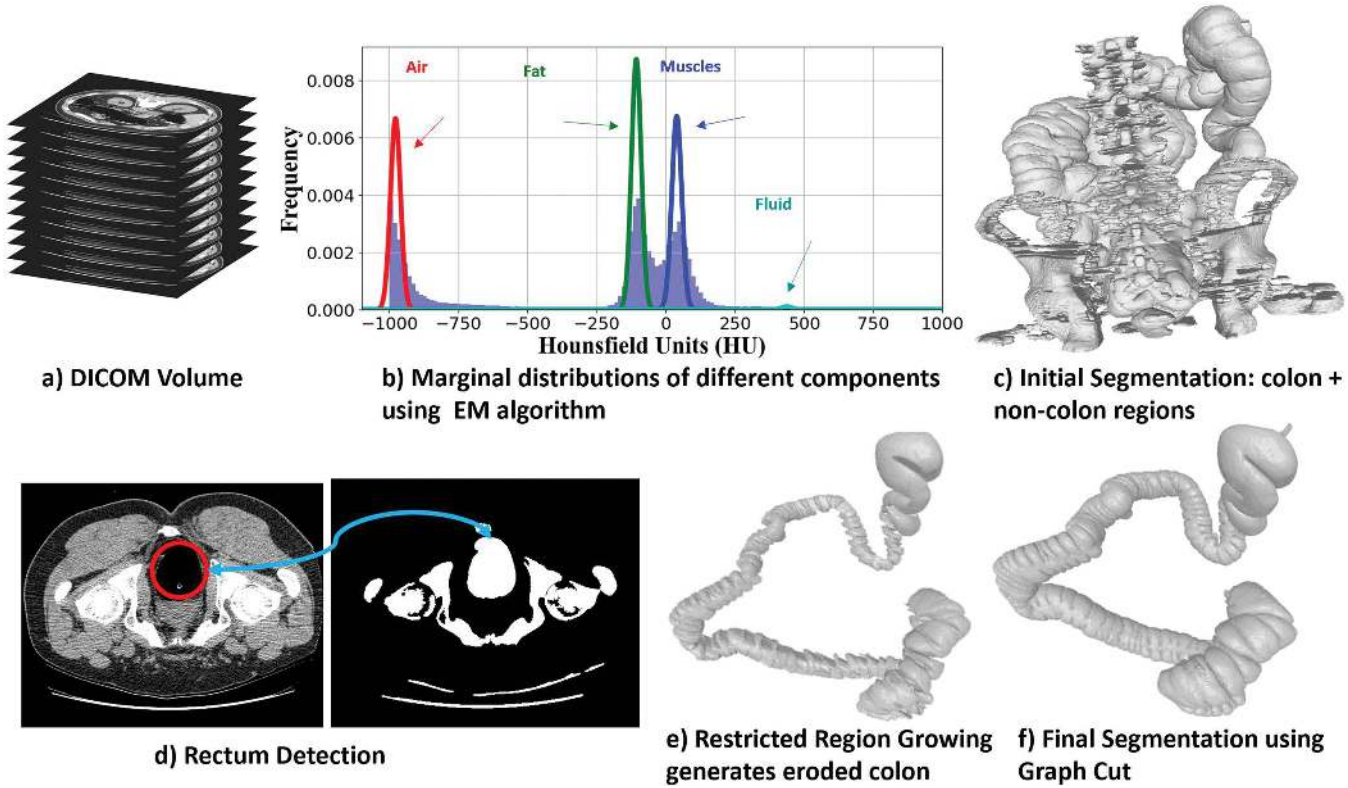

FIGURE 47.3 The classical approach for colon segmentation.

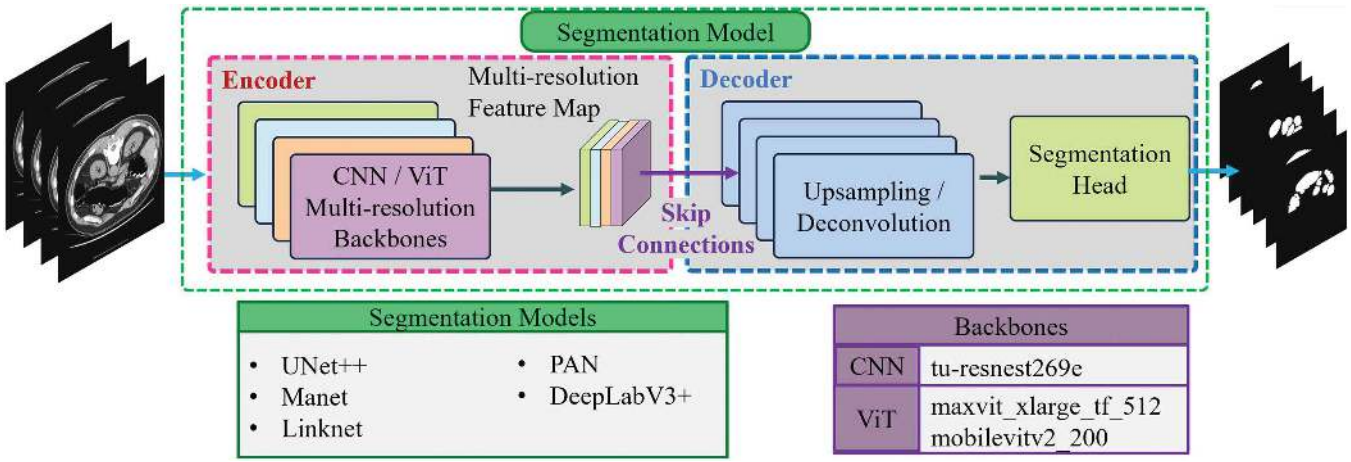

FIGURE 47.4 A meta-architecture consists of encoder and decoder. The encoder has a CNN-based backbone which generates multi-resolution feature maps that are passed to the decoder via skip connections. The decoders fuse these features through deconvolution blocks to generate a final feature map. The segmentation 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.

While most segmentation problems treat the objects to be segmented as a whole, that is, 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 of a 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. 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:

$$DiL = 1 - \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2}, \tag{47.3}$$

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 [31]. Further, $BL(x,b) = \frac{1}{N}\sum_{i=1}^{N} b_i \cdot x_i$ where 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 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 is the value in the boundary map at position i, that is the value in the distance map at the same position:

$$DM(M)_i = \min_{r \in \text{Boundary}(M)} \| r - i \|. \tag{47.4}$$

47.4 Results

We conducted a series of experiments to rigorously evaluate the effectiveness of the colon segmentation approaches. Unfortunately, most of the previous colon segmentation work, for example [12, 17] evaluated their models on either a private dataset or a public dataset whose annotation is still private. Therefore, we will assess our model performance on our private dataset.

The goal of the first experiment is to evaluate our MRF-based segmentation approach. We compared the predicted segmentation mask x to the ground truth y using the intersection over the union metric $\frac{y \cap x}{y \cup x}$. The results show that the accuracy of the MRF-based approach is 90% for 70% of the scans, the remaining 30% of scans need to be manually refined by radiologists.

In the second experiment, we investigate which state of the art (SOTA) segmentation architecture and backbone are optimum for the CNN-based model. Among the encoder-decoder architectures proposed in the literature, UNet [19], UNet++ [20], Linknet [21], FPN [22], PSPNet [23], PAN [24], DeepLabV3 [25], and DeepLabV3+ [26] are considered the most successful ones. The brute force experimentation with all 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, PSPNet especially 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: UNet++, Linknet, PAN, and 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 47.1 shows that UNet++ with resnest269e backbone gives the best performance for the base model.

In the third experiment, we conducted a patient-independent seven-fold cross-validation for the base model with resnest269e encoder and UNet++ decoder. To enrich the training dataset, we incorporate deformable augmentations, which is an appropriate augmentation in case of soft organs like colon. The total size of the CTC-Dataset is 42,609 CT images. In each fold, data is divided into 42 patients/7 patients, on average 36,469/6140 images, for training/validation, respectively. Table 47.2 indicates the validation metrics for the base model compared to other models. We employed different well-known metrics [32] that quantify the pixel-wise agreement between the predicted and ground truth segmentation such as the dice coefficient, Jaccard coefficient/IOU, Hausdorff distance, sensitivity (SN), specificity (SP), and so on.

In addition, it is important to highlight that we conducted a comparative analysis by contrasting our proposed method with the nnUNet [33], which is the SOTA 3D-based method for medical image segmentation. Due to the computation burden, the 3D network was evaluated with low-resolution images. However, the 2D-based proposed model performs as

TABLE 47.1

Validation Dice and IOU Scores for Multiple Architectures with Different Backbones

Architecture/Backbone	Dice Score	IOU
PAN/resnest269e	97.1	95.46
DeepLabV3+/resnest269e	97.35	95.74
Linknet/resnest269e	97.36	95.96
UNet++/resnest269e	97.44	95.75

TABLE 47.2

Evaluation of Volume Segmentations for nnUNet, Base Model (UNet++/resnest269e), C-LSTM, and the Proposed Model Using the Metrics Reported in [32]

Model	DICE	JACRD	bAVD	KAPPA	ICCORR	MUTINF	PROB	SN	SP
nnUNet*	98.79	97.61	0.0133	98.73	98.79	25.78	5E-05	98.85	99.94
Base Model	97.44	95.75	0.3136	97.39	97.47	22.72	1E-04	98.01	99.89
C-LSTM	89.16	80.68	0.39	88.62	91.10	21.30	5E-04	96.38	99.08

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

The dataset size for train/val: 29,430/7,786.

accurately as the 3D network without the need for down scaling the images and it has less computation complexity.

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 U-Net by replacing the convolutional layers in the encoder section with C-LSTM layers, as described in [34]. However, its performance is still lower than the base model, as shown in Table 47.2, 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.

The last experiment explores the effect of accurate segmentation on preserving colorectal polyp information. We segment six scans, which have ten polyps, using different segmentation approaches. These scans are from the test set and are not used in the training process. For each segmented colon, we investigated the presence of the polyps. We then assigned a score if a polyp is not affected, 0 if a polyp is missed and 1/2 if a polyp is partially affected. Finally, sensitivity (i.e., percentage of polyps even partially exist from the total polyps) and accuracy (i.e., the scores average) are calculated as shown in Table 47.3. The results show that unlike classical approach, DL-based approaches can preserve the polyp's information even partially, that is, sensitivity = 100%, which helps radiologists in data annotation.

TABLE 47.3

Evaluation of Effectiveness of Colon Segmentation on Polyp Detection

Metric	MRF-based	Base Model
Sensitivity	90	100
Accuracy	55	75

47.5 Discussion

In this chapter, we described an MRF- and a DL-based architecture for colon segmentation. We described a meta-architecture that was trained with CTC-Dataset consisting of 98 prone and supine scans. Through rigorous experiments and evaluations, we demonstrated the superiority of the CNN-based architecture, achieving an impressive validation Dice score of 98.68%. Additionally, a seven-fold cross-validation reinforced the consistency and robustness of our approach across different patient groups. Evaluation of preserved polyp information in segmented colons shows promising results in colon segmentation, setting a strong foundation for future applications in clinical settings.

47.6 Future Trends

This chapter presented two quantitative algorithms for colon segmentation in CTC: an MRF-based approach and a CNN-based method. The MRF-based approach formulates the problem as MAP estimation, minimizing an energy function to achieve spatial smoothness and data consistency, and employs a region-growing algorithm to refine the segmentation and exclude non-colonic tissues. On the other hand, the CNN-based approach leverages an encoder-decoder architecture with multi-resolution feature maps and custom per-pixel weighted dice loss, emphasizing boundary pixels to improve segmentation accuracy, especially in regions critical to detecting polyps.

Improved accuracy of colon segmentation using CNN identified by deep learning suggests further development by integrating advanced techniques such as few-shot learning (FSL) to overcome the limitations of traditional CNNs, which require large, densely annotated datasets. Potential enhancements include incorporating 3D contextual information through sequential episodic training (SET), guided by a Markov random field–based model for colon anatomy. Pairing this approach with contrastive learning is expected to further improve segmentation accuracy and polyp detection. The MRF-based approach may be further examined with additional scans and pathological CTC cases with various types of polyps. Improving the accuracy of lumen segmentation will enhance the accuracy of the front end of CTC.

REFERENCES

1. 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, p. 233–254, 2023.
2. "The American Cancer Society," [Online]. Available: https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/acs-recommendations.html.
3. R. L. Siegel, A. N. Giaquinto and A. Jemal, "Cancer statistics, 2024," *CA: A Cancer Journal for Clinicians,* vol. 74, no. 1, pp. 12–49, 2024.
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, 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,* vol. 20, p. 1–8, 2020.
6. S. Halligan, E. Dadswell, K. Wooldrage, J. Wardle, C. von Wagner, 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 multicentre randomised trials with economic evaluation (the SIGGAR trials).," *Health Technology Assessment (Winchester, England),* vol. 19, p. 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, p. 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.
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.
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 International Symposium on Biomedical Imaging: From Nano to Macro*, 2009.
11. Z. Ravindran, N. S. Das and others, "Automatic segmentation of colon using multilevel morphology and thesholding," in *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 2021.
12. K. Gayathri Devi, R. Radhakrishnan and others, "Automatic segmentation of colon in 3D CT images and removal of opacified fluid using cascade feed forward neural network," *Computational and Mathematical Methods in Medicine,* vol. 2015, 2015.
13. J. J. Nappi, A. H. Dachman, P. MacEneaney 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*. Vol. 4683. SPIE, 2002.
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, p. 205–219, 2018.
15. G. S. K. D. Z. A. Malhotra 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, p. 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 Hadish, "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, 2023.
19. 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*, 2015.
20. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging,* vol. 39, p. 1856–1867, 2019.
21. A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE Visual Communications and Image Processing (VCIP)*. IEEE, p. 1–4, 2017.
22. T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature pyramid networks for object detection," in *IEEE CVPR*, 2017.
23. H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid scene parsing network," in *IEEE CVPR*, 2017.
24. H. Li, P. Xiong, J. An and L. Wang, "Pyramid attention network for semantic segmentation," *arXiv preprint arXiv:1805.10180,* 2018.
25. L. Chen, G. Papandreou, F. Schroff and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587,* 2017.
26. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
27. P. Wesp, S. Grosu, A. Graser, S. Maurus, C. Schulz, T. Knösel, M. P. Fabritius, B. Schachtner, B. M. Yeh, C. C. Cyran and others, "Deep learning in CT colonography: Differentiating premalignant from benign colorectal polyps," *European Radiology,* vol. 32, p. 4749–4759, 2022.
28. I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101,* 2017.
29. I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983,* 2016.
30. P. Iakubovskii, *Segmentation Models Pytorch,* GitHub, 2019.
31. 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)*.
32. A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Medical Imaging,* vol. 15, p. 1–28, 2015.
33. F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods,* vol. 18, p. 203–211, 2021.
34. A. Arbelle and T. R. Raviv, "Microscopy cell segmentation via convolutional LSTM networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, p. 1008–1012, 2019.