Metadata of the chapter that will be visualized in SpringerLink

Book Title	Pattern Recognition	
Series Title		
Chapter Title	Colon Segmentation Using Guided Sequential Episodic Training and Contrastive Learning	
Copyright Year	2025	
Copyright HolderName	The Author(s), under exclusive license to Springer Nature Switzerland AG	
Corresponding Author	Family Name	**Harb**
	Particle	
	Given Name	**Samir**
	Prefix	
	Suffix	
	Role	
	Division	Computer Vision and Image Processing Laboratory (CVIP)
	Organization	University of Louisville
	Address	Louisville, KY, USA
	Division	
	Organization	Higher Technological Institute
	Address	10th of Ramadan City, Egypt
	Email	safara01@louisville.edu
Author	Family Name	**Ali**
	Particle	
	Given Name	**Asem**
	Prefix	
	Suffix	
	Role	
	Division	Computer Vision and Image Processing Laboratory (CVIP)
	Organization	University of Louisville
	Address	Louisville, KY, USA
	Email	
Author	Family Name	**Yousuf**
	Particle	
	Given Name	**Mohamed**
	Prefix	
	Suffix	
	Role	
	Division	Computer Vision and Image Processing Laboratory (CVIP)
	Organization	University of Louisville
	Address	Louisville, KY, USA
	Division	Faculty of Engineering
	Organization	Ain Shams University
	Address	Cairo, Egypt

	Email	
Author	Family Name	**Elshazly**
	Particle	
	Given Name	**Salwa**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Kentucky Imaging Technologies
	Address	Louisville, KY, USA
	Email	
Author	Family Name	**Farag**
	Particle	
	Given Name	**Aly**
	Prefix	
	Suffix	
	Role	
	Division	Computer Vision and Image Processing Laboratory (CVIP)
	Organization	University of Louisville
	Address	Louisville, KY, USA
	Email	

Abstract	Accurate colon segmentation on abdominal CT scans is crucial for various clinical applications. In this work, we propose an accurate approach to colon segmentation from abdomen CT scans. Our architecture incorporates 3D contextual information via sequential episodic training (SET). In each episode, we used two consecutive slices, in a CT scan, as support and query samples in addition to other slices that did not include colon regions as negative samples. Choosing consecutive slices is a proper assumption for support and query samples, as the anatomy of the body does not have abrupt changes. Unlike traditional few-shot segmentation (FSS) approaches, we use the episodic training strategy in a supervised manner. In addition, to improve the discriminability of the learned features of the model, an embedding space is developed using contrastive learning. To guide the contrastive learning process, we use an initial labeling that is generated by a Markov random field (MRF)-based approach. Finally, in the inference phase, we first detect the rectum, which can be accurately extracted using the MRF-based approach, and then apply the SET on the remaining slices. Experiments on our private dataset of 98 CT scans and a public dataset of 30 CT scans illustrate that the proposed FSS model achieves a remarkable validation dice coefficient (DC) of 97.3% (Jaccard index, JD 94. 5%) compared to the classical FSS approaches 82.1% (JD 70.3%). Our findings highlight the efficacy of sequential episodic training in accurate 3D medical imaging segmentation. The codes for the proposed models are available at https://github.com/Samir-Farag/ICPR2024.

Keywords (separated by '-')	Colon segmentation - Deep learning - Few-shot

Colon Segmentation Using Guided Sequential Episodic Training and Contrastive Learning

Samir Harb[1,2(✉)], Asem Ali[1], Mohamed Yousuf[1,3], Salwa Elshazly[4], and Aly Farag[1]

[1] Computer Vision and Image Processing Laboratory (CVIP), University of Louisville, Louisville, KY, USA
safara01@louisville.edu
[2] Higher Technological Institute, 10th of Ramadan City, Egypt
[3] Faculty of Engineering, Ain Shams University, Cairo, Egypt
[4] Kentucky Imaging Technologies, Louisville, KY, USA

Abstract. Accurate colon segmentation on abdominal CT scans is crucial for various clinical applications. In this work, we propose an accurate approach to colon segmentation from abdomen CT scans. Our architecture incorporates 3D contextual information via sequential episodic training (SET). In each episode, we used two consecutive slices, in a CT scan, as support and query samples in addition to other slices that did not include colon regions as negative samples. Choosing consecutive slices is a proper assumption for support and query samples, as the anatomy of the body does not have abrupt changes. Unlike traditional few-shot segmentation (FSS) approaches, we use the episodic training strategy in a supervised manner. In addition, to improve the discriminability of the learned features of the model, an embedding space is developed using contrastive learning. To guide the contrastive learning process, we use an initial labeling that is generated by a Markov random field (MRF)-based approach. Finally, in the inference phase, we first detect the rectum, which can be accurately extracted using the MRF-based approach, and then apply the SET on the remaining slices. Experiments on our private dataset of 98 CT scans and a public dataset of 30 CT scans illustrate that the proposed FSS model achieves a remarkable validation dice coefficient (DC) of 97.3% (Jaccard index, JD 94. 5%) compared to the classical FSS approaches 82.1% (JD 70.3%). Our findings highlight the efficacy of sequential episodic training in accurate 3D medical imaging segmentation. The codes for the proposed models are available at https://github.com/Samir-Farag/ICPR2024.

Keywords: Colon segmentation · Deep learning · Few-shot

This work has been funded by NSF grant 2124316.

A. Antonacopoulos et al. (Eds.): ICPR 2024, LNCS 15313, pp. 1–16, 2024.
https://doi.org/10.1007/978-3-031-78201-5_5

1 Introduction

Automated image segmentation plays a crucial role in medical imaging research and clinical applications by automating or facilitating the delineation of anatomical structures and other regions of interest. The segmentation step is of significant importance to facilitate accurate identification and delineation of structures or abnormalities for clinical applications such as lesion localization, disease diagnosis, and prognosis [30,31]. Specifically, automatic colon segmentation is a key step for medical image analysis pipelines (e.g. colonography [3,16]), because any inaccuracies at the segmentation stage will carry through to subsequent steps. This underscores the importance of prioritizing the segmentation process and improving its effectiveness, which consequently leads to performance enhancements in the next stages of this pipeline.

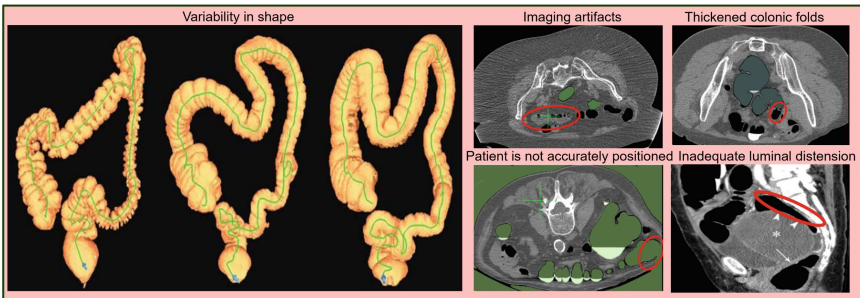

Fig. 1. Examples of challenges that hinder accurate segmentation of the colon [3,11].

However, segmenting the colon regions accurately from abdominal CT scans poses significant challenges, as depicted in Fig. 1. First, colon regions exhibit highly variable and asymmetric topology [22], and their positions vary between different CT images [9]. Second, distinguishing colon regions 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, e.g., bones. Lastly, patient preparation imperfections, such as residual stool and lesions, can lead to disjointed colon segments. These complexities inherent in colon segmentation, particularly in scenarios where automated algorithms are indispensable, may confuse segmentation algorithms [9]

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 MRF-based models, e.g. [3,5,21,26], edge detection, region growth and division, e.g. [6,7,17,20,21], or hybrid segmentation algorithms [14]; and (2) deep learning (DL) approaches, e.g. [2,10,15,30], which exploit the available data to learn complicated high-level characteristics

that can be used for segmentation, unlike the classical approaches that focus on low-level traits, which may not be as helpful for segmentation.

Although DL approaches are successful segmentation tools, Jakob Wasserthal et al. [30], who developed the state-of-the-art (SOTA) segmentation tool, Totalsegmentator, reported that the colon posed the most significant challenges, with a failure rate of ∼35% of cases. This failure was mainly attributed to difficulties in precisely segmenting the subtle details of the colon.

Traditional deep Convolutional Neural Networks (CNNs) adept at semantic segmentation often encounter challenges, relying on a plethora of densely annotated images for effective training and struggling to generalize to unfamiliar object classes. This issue is exacerbated in medical imaging, where the dearth of annotations hampers the applicability of conventional methods. Recently, few-shot learning (FSL) has emerged as a prominent deep learning approach to equip a model with the ability to segment unseen semantic classes by learning from just a few labeled images of this unseen class during inference, without necessitating model retraining. Hence, Few-Shot Segmentation (FSS) was introduced to address the challenges of medical image segmentation by leveraging knowledge distilled from labeled samples (support) to segment unlabeled samples (query). FSS learns tasks composed of base class in an episodic training manner and segments unseen classes in the form of tasks in the inference stage.

One of the pioneering DL networks proposed to utilize FSL in natural images is PANet [29]. Prior to PANet, FSS methods demonstrated unsatisfactory generalization due to a lack of separation between knowledge extraction and segmentation processes, as well as the utilization of support data solely for masking purposes. PANet addressed these issues by introducing a separation between prototype extraction (which involves feature extraction from support images and subsequent prototype extraction from these features, along with feature extraction from query images) and non-parametric metric learning (which segments the query image by computing the cosine distance between each support class prototype and query features at each spatial location). Furthermore, PANet uses annotations to supervise Few-Shot Learning. To eliminate the need for annotations during training, Ouyang, Cheng et al. [19] developed a self-supervised FSS framework, SSL-ALPNet, that exclusively utilizes superpixel-based pseudo-labels for supervision. In addition, an adaptive local prototype module is presented to mitigate the challenge of foreground-background imbalance in medical image segmentation. Wu, Huisi et al. [31] proposed AAS-DCL to combine dual contrastive learning and anatomical guidance to enhance feature discriminability and data utilization to help few-shot medical image segmentation.

In this work, we propose a novel FSS approach for precise colon segmentation in abdominal CT scans, addressing the inherent challenges of this critical medical imaging task. Our proposed approach introduces an episodic segmentation strategy that takes advantage of sequential episode training and contrastive learning techniques. Unlike traditional few-shot segmentation approaches, our method employs supervised episodic training, facilitating enhanced feature discriminability and segmentation accuracy. In particular, we incorporate unrelated

slices rich in anatomical structures to provide vital background guidance, further refining the segmentation process. Based on the AAS-DCL framework [31], our approach integrates dual contrastive learning (DCL) and anatomical guidance, culminating in improved feature extraction and segmentation performance. In addition, we introduce a novel MRF-based rectum detection and initial labeling technique, enhancing the robustness and accuracy of the proposed approach. The primary contributions of our work are as follows:

i) Develop an MRF-based rectum detection and initial labeling method, contributing to improved accuracy and robustness of the overall segmentation process.
ii) Integrate supervised sequential episodic training and contrastive learning techniques to enhance feature discriminability and segmentation accuracy, while incorporating unrelated slices rich in anatomical structures to provide essential background guidance.
iii) Enhance feature extraction and segmentation performance through the integration of dual contrastive learning with anatomical guidance.

2 Method

Our approach aims to accurately segment the colon in abdominal CT scans. We use a method that combines 3D information (through SET) with 2D segmentation models. This allows us to avoid the high computational costs of complex 3D neural networks while still achieving precise results. The 2D models are efficient and flexible, handling individual CT images well even with irregular sampling.

2.1 Proposed Episodic Segmentation Approach

The traditional episodic training strategy for the few-shot segmentation (FSS) approach involves training a model over a large number of epochs, with multiple episodes in each epoch. So, a dataset, in episodic training, is arranged into multiple episodes and each episode consists of support and query pairs. For a set of images $\mathcal{X}$ and its corresponding set of binary masks $\mathcal{Y}$, we define the support set $\mathcal{S} = \{x_s^c, y_s^c\}$ and the query set $\mathcal{Q} = \{x_q^c, y_q^c\}$, where $x_{s(q)}^c \in \mathcal{X}$, $y_{s(q)}^c \in \mathcal{Y}$, and the superscript c represents an arbitrary class in a set of classes $\mathcal{C}$. Since few-shot segmentation approaches were introduced to take advantage of distilled knowledge from labeled samples for segmenting unlabeled ones, in these approaches, a model is trained to identify a set of classes $\mathcal{C}_{tr}$ in a training dataset $\mathcal{D}_{tr}$. But it never sees the set of classes $\mathcal{C}_{ts}$ in the test dataset $\mathcal{D}_{ts}$. Then, during the inference, the model is used to segment the unseen classes $\mathcal{C}_{ts}$ in $\mathcal{D}_{ts}$ using annotated samples of these classes, without the need to re-train the model.

We propose an FSS-like approach in which we use support and query sets, but unlike the classical FSS approaches, we use the episodic training strategy in a supervised manner. Therefore, training and test classes are the same, that is, $\mathcal{C}_{tr} = \mathcal{C}_{ts} = \{colon\}$, but $\mathcal{D}_{tr} \neq \mathcal{D}_{ts}$ where $\mathcal{D}_{tr}$ contains training scans and

$\mathcal{D}_{ts}$ contains testing scans. Moreover, to enhance the discriminability of learned features of the model, we exploit unrelated slices $U = \{x_u^{\bar{c}}, y_u^{\bar{c}}\}$ but rich with anatomical structures (i.e., $\bar{c}$ non-colon regions, e.g., liver) to provide more background guidance. Using support, query, and unrelated features, extracted from x_s^c, x_q^c, and $x_u^{\bar{c}}$, respectively, we develop an embedding space using contrastive learning to pull closer (x_s^c, x_q^c) and push farther $(x_q^c, x_u^{\bar{c}})$.

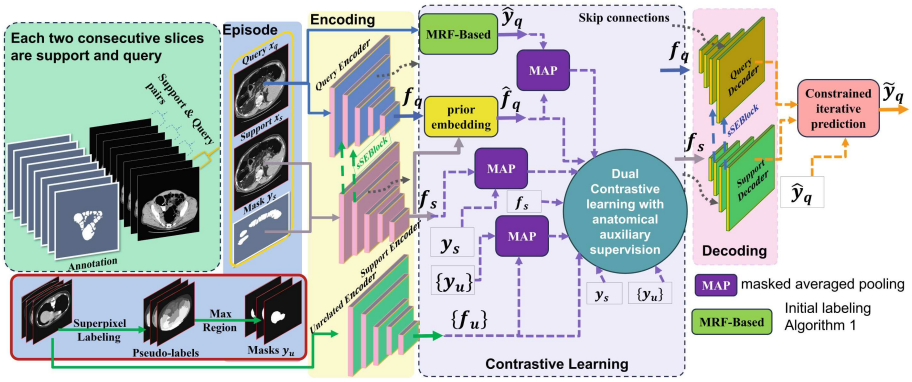

Fig. 2. The proposed SET FSS approach with DCL. An MRF-based auxiliary supervision is used to enhance the baseline AAS-DCL. The workflow starts by arranging scans into pairs of consecutive slices (x_s^c, x_q^c). Then unrelated slices are selected as negative samples $\{x_u^{\bar{c}}\}$. The episode $(x_s^c, x_q^c, \{x_u^{\bar{c}}\})$ is fed into encoders to extract features $(f_s, f_q, \{f_u\})$. Masked average pooling is applied on these features and contrastive learning is used to generate an embedding space. Finally, decoders with skip connection are used to estimate the final segmentation $\widetilde{y}_q$, which is iteratively refined using initial labeling.

We build on the AAS-DCL approach [31], as shown in Fig. 2, to combine DCL and anatomical guidance to enhance feature discriminability. However, unlike the AAS-DCL approach, we use SET, in which support and query samples are consecutive slices in a CT scan. The motivation behind this is that the anatomy of the human body does not have abrupt changes 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. Therefore, using consecutive slices as support and query samples simplifies the segmentation approach. Moreover, for more guidance, we start with an initial labeling that is generated using an MRF-based approach (Algorithm 1). This enhances the DCL. Finally, in the inference stage, we first detect the rectum, which can be accurately extracted as shown in Fig. 4, then apply the sequential episodic approach.

Unrelated Slices Selection: The proposed workflow starts by arranging CT scans into pairs of consecutive slices (x_s^c, x_q^c). Then, in each episode, we randomly

select three unrelated slices as negative samples $\{x_u^{\bar{c}}\}$. These unrelated slices do not include colon regions but they may have irrelevant organs or tissues. To define masks $\{y_u^{\bar{c}}\}$ for unrelated samples $\{x_u^{\bar{c}}\}$, we employ an unsupervised graph cut-based algorithm [8], offline, which generates superpixel segmentation. These pseudo-labels are binarized by choosing the dominant superpixel (i.e., the largest connected region) in each pseudo-label as a target and other superpixels as a background. Then an encoder is used to extract the features $\{f_u^{\bar{c}}\}$ from $\{x_u^{\bar{c}}\}$. Finally, these features and their masks $\{f_u^{\bar{c}}, y_u^{\bar{c}}\}$ are included in the AAS-DCL scheme.

Dual Contrastive Learning: To provide more background guidance, we exploit the unrelated slices with query and support slices in contrastive learning. Inspired by the baseline AAS-DCL approach [31], we combine prototypical contrastive learning and contextual contrastive learning to form a DCL scheme, which makes the features of the colon regions closer to other characteristics of dissimilar tissues. The infoNCE loss [18] $\mathcal{L}(v_q, v_s, v_u)$ is used for the training process of the contrastive learning module.

$$\mathcal{L}(v_q, v_s, v_u) = -v_q.v_s/\tau + \log \sum_{i=1}^{n} \exp(v_q.v_{ui}/\tau),$$

where τ is a control parameter, n is the number of negative samples, v_q, v_s, and v_u are the query, support and background prototypes, respectively. These prototypes are generated by the global average pooling of features and the corresponding masks. However, these prototypes cannot acquire intra-class variations. To overcome this problem, patch-based prototypes may be used.

Prototypical Contrastive Learning: Prototype-based learning is based on the generation of prototypes that discriminate between the features of the foreground and the background. In this approach, support features $\{f_s\}$ and their corresponding masks $\{y_s\}$ are used to generate the colon prototype using the masked averaged pooling (MAP) operation [34] $v_s = \frac{\sum_r y_s(r).f_s(r)}{\sum_r y_s(r)}$.

Unlike the baseline AAS-DCL approach, which uses global average pooling to calculate the query prototype, we exploit the initial query mask $\hat{y}_q$ to calculate the query prototype using masked average pooling. Also, instead of using the query feature f_q, we employ a prior embedding module [31] to enhance the query feature. The enhanced query feature $\hat{f}_q$ further activates the foreground information in the query prototype $v_q = \frac{\sum_r \hat{y}_q(r).\hat{f}_q(r)}{\sum_r \hat{y}_q(r)}$.

Similarly, unrelated features $\{f_u\}$ and their corresponding masks $\{y_u\}$ are also used to generate the background prototype v_u using masked averaged pooling.

To overcome intra-class variations and to exploit information about other structures around colon regions as unrelated samples, we employed patch-based learning [19]. In this method, the support feature and its mask are divided into

patches, then these patches are used to generate a colon prototype and a background prototype depending on a threshold. This scheme increases the number of negative samples and distinguishes between the local characteristics of different tissues. Again, unlike the baseline AAS-DCL approach, we exploit the initial mask of the query $\hat{y}_q$ to calculate the query prototype using the masked average pooling.

Contextual Contrastive Learning: Finally, to guide feature maps focusing on rich contextual information, a spatial attention block [24] is employed to process the support feature f_s, enhanced query feature $\hat{f}_q$ and unrelated features $\{f_u\}$. Then the processed features are averaged and used in contextual contrastive learning, for more details see [31].

Iterative Prediction: For accurate segmentation, iterative optimization methods [28,32] are used to combine the prediction of the query with the query feature by convolution. Unlike the baseline [31], we guide the iterative process using the initial labeling to promote the fusion of the query feature and the predicted mask. The query prediction is updated through the similarity consistency constraint, in which we also use initial labeling to calculate a similarity map between support and query features.

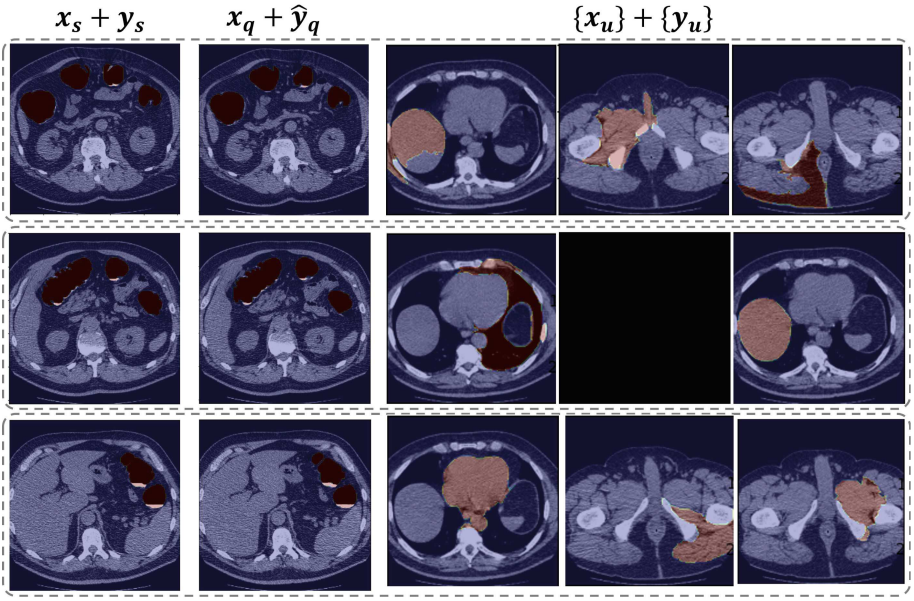

Fig. 3. Examples of episodes from training dataset. Each row represents a single episode that includes labeled support, query with initial labeling, and unrelated labeled slices.

Training Stage: In this stage, two consecutive slices are randomly selected as a pair of support and query. In addition, three randomly selected unrelated slices are added to this pair to form an episode, as shown in Fig. 3. Each episode is fed into the encoder-decoder sSENet [25] for feature extraction and reconstruction. Then the cross-entropy loss uses the prediction of this module to calculate a prediction error against the ground truth. The prediction error and contrastive learning loss, which are computed using the extracted features and the initial query mask, are backpropagated to train the network.

Inference Stage: This stage starts by detecting the rectum and the initial mask for a given abdomen CT scan using the proposed MRF-based approach. Subsequently, a rectum slice is considered a support sample and the consecutive slice is a query. In addition, three randomly selected unrelated slices are added to this pair to form an episode. Each episode is fed into the trained model to generate the prediction of the query. Then, the segmented query slice will be the support sample for the consecutive slice in the sequence. This iterative process continues until all colon regions are successfully segmented.

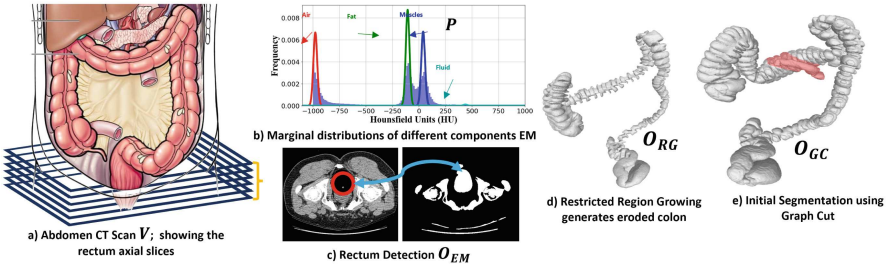

Fig. 4. MRF-based initial labeling approach. Rectum is the only region, in the lower CT slices that has air, and it can be easily identified as a disk-like region that has low Hounsfield. First, EM is used to calculate the empirical distributions $\mathbf{P}$ of Hounsfield intensities in a DICOM volume $\mathbf{V}$. Thresholds between air and fat and between muscle and fluid are used to generate $\mathbf{O}_{EM}$. RG algorithm is applied starting from the rectum to generate an eroded colon $\mathbf{O}_{RG}$. This guarantees that other organs, e.g., small intestine, are not merged with $\mathbf{O}_{RG}$. Finally, $\mathbf{O}_{RG}$ is refined through an optimization technique to generate $\mathbf{O}_{GC}$, which still may have other structures (colored) misclassified as colon. (Color figure online)

2.2 MRF-Based Rectum Detection and Initial Labeling

To generate an initial labeling, we develop a multi-step approach, which employs three algorithms: Expectation-Maximization (EM) to calculate the empirical distributions of Hounsfield units (HU) in a DICOM volume, Region Growing

(RG) to generate initial labeling by identifying colon regions starting from the rectum, and Graph Cut (GC) to estimate the initial mask of colon regions. The main components of a colon are the air, for which the characteristic peaks are almost at -1000 HU [17], and the opacified fluid whose Hounsfield intensity is greater than 300 HU. To extract the colon components, first we estimate the marginal densities of air, fat, muscle, and fluid, in an abdomen scan, by fitting four Gaussian components using the EM algorithm, as shown in Fig. 4-b. Then, we identify colon regions using two thresholds. As shown in Fig. 4-c, the rectum is the only region, in the lower CT slices of an abdomen scan, that has air. Therefore, the rectum region can be easily identified as a disk-like region with a low Hounsfield unit. This region is used as a starting seed, from which other colon regions are extracted by the proposed model.

The problem is formulated as the maximum-A posterior estimate of an MRF model, which involves finding the labeling that minimizes the following energy function $E(\tilde{y})$ (Eq. (1)) that combines both the spatial smoothness and data consistency.

$$E(\tilde{y}) = \sum_{\{r,t\}\in\mathcal{N}} V(\tilde{y}_r, \tilde{y}_t) + \sum_{r\in\mathcal{P}} D(\tilde{y}_r), \tag{1}$$

where $\mathcal{N}$ represents the set of neighboring pixel pairs (r, t), $V(.,.)$ is the potential function that penalizes label inconsistencies between neighboring pixels, and $D(.)$ is the data penalty term that measures how well the labeling $\tilde{y}_r$ matches the observed data. The minimization of the energy function in Eq. (1) using a graph cut generates the initial labeling result. The Algorithm 1 summarizes this approach.

Algorithm 1. MRF-based segmentation approach

0: **Input**: DICOM volume **V**
1: Calculate the histogram **P** of HU values in **V**
2: Apply EM algorithm, and identify air and fluid regions O_{EM}
3: Detect rectum region in O_{EM}
4: Starting from rectum region, apply RG algorithm to extract colon O_{RG} from O_{EM}
5: Use O_{RG} as a seed for GC and minimize $E(\tilde{y})$ to extract initial labeling O_{GC}

3 Experiments

Dataset: We conducted experiments on our private dataset having abdominal CT scans of 49 patients in both supine and prone positions. Experts annotated the colon segments in these 98 CT scans. Also, for the sake of comparison, we use the synapse public dataset [12] which has been used by several SOTA approaches. In our work, we refer to this dataset as SABS. It contains 30 abdominal CT scans. In SABS dataset, 13 organs were manually annotated (colon is not included) by 2 experienced undergraduate students and verified by a radiologist [1]. From

these two datasets, we created three training datasets, $\mathcal{D}_{tr_1}$, $\mathcal{D}_{tr_2}$ and $\mathcal{D}_{tr_3}$, and a testing dataset $\mathcal{D}_{ts}$:

i $\mathcal{D}_{tr_1}$ **(SABS):** Consists of 30 scans from the SABS CT dataset, with annotated organs labeled 1 through 13. Specifically, the labels include spleen (1), right kidney (2), left kidney (3), and so forth, up to the left adrenal gland (13).
ii $\mathcal{D}_{tr_2}$ **(SABS + CTC68):** Combines the SABS dataset (with 13 annotated organs) and 68 scans (34 prone and 34 supine) from our private dataset (with annotated colon). Consequently, the combined dataset covers spleen (1), right kidney (2), left kidney (3), and so forth, up to the left adrenal gland (13), and includes the colon as label 14.
iii $\mathcal{D}_{tr_3}$ **(CTC68):** Includes 68 scans (34 prone and 34 supine) from our private dataset (with annotated colon).
iv $\mathcal{D}_{ts}$ **(CTC30):** Encompasses 30 scans (15 prone and 15 supine) from our private dataset (with annotated colon).

Evaluation Metrics. We employed both DC and JD to quantify the pixel-wise agreement between the predicted and ground truth segmentation [27]. This dual assessment approach considers both the overlap and spatial agreement between the predicted and ground truth colon regions.

Technical and Implementation Details: We implemented our framework with PyTorch, based on the official baseline implementation https://github.com/cvszusparkle/AAS-DCL_FSS, on a Nvidia TITAN RTX with 24 GB. Among the different available off-the-shelf fully convolutional networks, we utilized ResNet101 that guaranteed high spatial resolutions in feature maps. As a preprocessing step, we first resize the 2D slices to 256×256 resolution and divide data into 4 patches for prototypical contrastive learning. Our proposed SET starts with a learning rate of 10^{-4}, a batch size of 1, and applies polynomial decay. Adam optimization with power = 0.95 and weight decay = 10^{-7} is used over 100 epochs. Data augmentation includes random adjustments to sharpness and lightness. For high-resolution feature maps, a fully convolutional ResNet101 pre-trained on MS-COCO processes 256×256 images to $256 \times 32 \times 32$ maps. Training uses a Local Pooling Window of 4×4, reducing to 2×2 for inference. Training on a Nvidia TITAN RTX GPU takes 3 h, using 3 GB memory, on average for the proposed model.

Standard FSS Approaches vs the Proposed SET FSS Approach: Since the proposed approach uses the FSS concept of support and query sets, we compare its performance against standard FSS approaches. To highlight the high performance of our proposed SET FSS approach with respect to the standard FSS segmentation approaches, we evaluated the SSL-ALPNet [19] model and the AAS-DCL [31] network, which is our baseline, in the colon segmentation problem. The experimental results on the test set $\mathcal{D}_{ts}$, shown in Table 1, shed light on the performance of various model configurations in colon segmentation.

Table 1. Comparison of validation DC and JD on $\mathcal{D}_{ts}$ dataset for our proposed models against the SOTA models, with different training and initialization settings.

	Model	Training	Initialization	DC	JD
SOTA	SSL-ALPNet [19]	SABS	None	34.1%	21.0%
		SABS + CTC	None	81.7%	70.0%
		CTC	None	82.1%	70.3%
	AAS-DCL [31]	SABS	None	16.0%	8.8%
		SABS + CTC	None	65.5%	49.5%
		CTC	None	68.8%	53.2%
Proposed	Guided-AAS-DCL	SABS	Superpixel	61.0%	44.2%
		SABS + CTC	Superpixel+MRF	83.0%	71%
		CTC	MRF	96.3%	92.9%
	SET-DCL	CTC	None	96.8%	93.7%
	Guided-SET-DCL	**CTC**	**MRF**	**97.3%**	**94.5%**

First, we used the standard FSS technique, in which we train SSL-ALPNet and AAS-DCL models using $\mathcal{D}tr1$ (i.e., self-supervised learning by training networks with data that included superpixel results instead of annotations). As expected, standard FSS techniques do not perform well in this scenario. This is due to many reasons, such as uncertainties in the dataset (e.g., prep deficits, patient conditions, and scanner settings and errors). In addition, the learned embedding space of the prototypes of different organs in the $\mathcal{D}tr1$ dataset has different distributions than the colon prototype due to the characteristics of different tissues. Specifically, SSL-ALPNet trained in SABS achieved 34.1% DC and 21.0% JD, and AAS-DCL trained on $\mathcal{D}tr1$ achieved 16.0% DC and 8.8% JD. For learning a more general embedding space, in the second experiment, we included the colon in the training phase. So, we used $\mathcal{D}tr2$ to train the two models (i.e., supervised learning by training networks with data including the colon along with the other 13 organs). This drastically enhances the performance of the models. SSL-ALPNet trained on $\mathcal{D}tr2$ achieved 81.7% DC and 70.0% JD, while AAS-DCL trained on $\mathcal{D}tr2$ achieved 65.5% DC and 49.5% JD.

To explore the upper limit of the performance of the models, we used the purely supervised learning scheme by training the models using $\mathcal{D}tr3$. The SSL-ALPNet trained model provides decent performance, achieving 82.1% DC and 70.3% JD, because the SSL-ALPNet model ensures that each prototype exclusively represents a distinct part of the object-of-interest. This enables precise localization of colon structures by preserving intricate local details crucial to segmentation accuracy. However, the AAS-DCL model needs more guidance to enhance its performance, achieving only 68.8% DC and 53.2% JD.

Ablation Study: The proposed approach depends on the initial labeling and sequential episodic learning. Table 1 summarize effects of these components. In order to enhance the performance of the baseline model, we guide the DCL

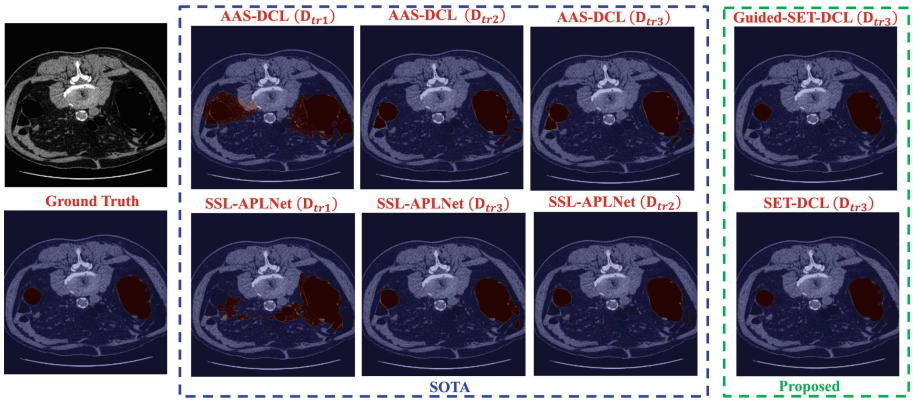

Fig. 5. Qualitative results on different training settings show that the results of SOTA FSS approaches include artifacts, on the other hand, the proposed method achieves desirable segmentation results that are close to ground truth.

using an initial labeling as explained in the proposed approach. Also, we add the constraint on a query slice to be within 5 neighbors from the support slice. This limits the changes in the colon structure. The first Guided-AAS-DCL model is trained using $\mathcal{D}tr1$ and the initial labeling for the organ of interest is estimated using the superpixel approach. The initialization and the neighbor constraint enhance the model performance from 16.0% to 61.0% DC and from 8.8% to 44.2% JD. Adding colon scans with MRF-based initial labeling to the training dataset in $\mathcal{D}tr2$ boosts the model performance, yielding a DC of 83.0% and a JD of 71%. Finally, the supervised learning performance of the Guided-AAS-DCL model reaches 96.3% DC and 92.9% JD. This highlights that the synergistic fusion of initial labeling and query constraint promises to deliver precise and reliable colon segmentation results.

Exploiting the anatomical structure of the colon, we propose the sequential episodic training SET-DCL FSS approach, in which the support and query are neighboring slices. Additionally, the inference phase starts with the detected rectum slices as a support and then sequentially segments the remaining slices where each segmented slice acts as a support slice for the consecutive query slice in the CT scan. Without any additional initialization, the proposed SET-DCL model exhibits a DC of 96.8% and a JD of 93.7%, better than the Guided-AAS-DCL model. Moreover, leveraging MRF-based initialization further enhances the performance of the proposed Guided-SET-DCL model's performance further, resulting in a remarkable DC of 97.3% and a JD of 94.5%. This underscores the efficacy of MRF-based initialization and sequential episodic training in increasing segmentation accuracy.

Figures 5 and 6 show the robustness of the proposed framework that consistently produces satisfactory results, especially for training solely with the CTC. Also, Fig. 7 provides more illustrations on how the proposed approach accurately

segments colon parts, while other SOTA approaches may miss parts and have some artifacts.

Supervised Learning Scheme: Finally, since our proposed approach depends on supervised learning, to compare against the SOTA CNN-based encoder-decoder segmentation architectures trained using a supervised learning scheme, we trained the PAN model [13] that is paired with resnest269e [33] backbone and U-Net model [23] using $\mathcal{D}tr3$ then we tested them on $\mathcal{D}_{ts}$. 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 scans. Therefore, we explore the fusion of C-LSTM with U-Net by replacing the convolutional layers in the encoder section with C-LSTM layers [4]. As shown in Table 2, our proposed approach outperforms the SOTA approaches. Specifically, the DC for our proposed approach (Guided-SET-DCL) is 97.3%, which is higher than MRF-based (87.9%), C-LSTMs (89.2%), U-Net (85.0%), and PAN (97.1%). Similarly, the JD for our proposed approach is 94.5%, which also outperforms MRF-based (84.5%), C-LSTMs (80.7%), U-Net (80.0%), and PAN (95.5%). The C-LSTM has a lower performance 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.

Although the experiments were conducted to segment the colon, we believe that the same approach can be successfully used to segment other organs that are scanned as sequential slices that do not have abrupt changes.

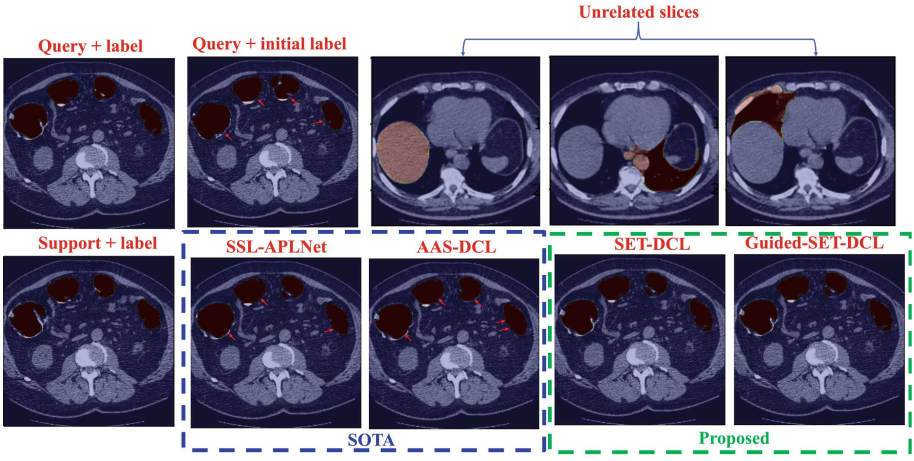

Fig. 6. An example of an episode: query with ground truth, query with initial labeling, unrelated slices with labels, and support with label. The qualitative results show that SOTA FSS approaches miss colon semilunar folds (shown in red arrows); on the other hand, the proposed method achieves desirable segmentation results that are close to ground truth. (Color figure online)

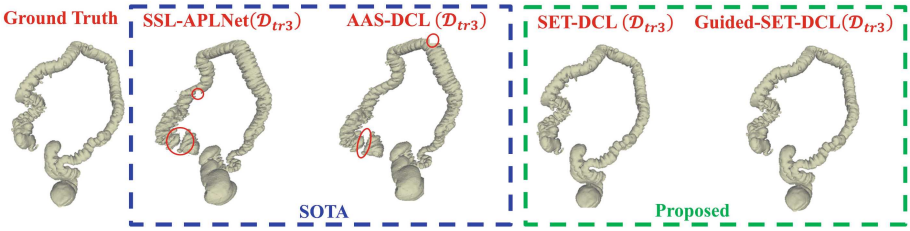

Fig. 7. Ground truth 3D colon and results of the proposed method compared to the SOTA FSS approaches. The qualitative results show that the SOTA FSS approaches miss parts and generate incomplete colon, on the other hand, the proposed method generates accurate colon segments.

Table 2. Comparison of validation DC and JD on $\mathcal{D}_{ts}$ dataset for our proposed approach against CNN-based SOTA architectures.

Metric	MRF-based	C-LSTMs [4]	U-Net [23]	PAN [13]	Guided-SET-DCL
DC	87.9%	89.2%	85.0%	97.1%	**97.3%**
JD	84.5%	80.7%	80.0%	**95.5%**	94.5%

4 Conclusions

We proposed an FSS approach that addresses the significant challenge of accurate colon segmentation in abdominal CT scans. Through the integration of a classical segmentation model, i.e., MRF model, deep learning, and sequential episodic training, we developed a comprehensive approach for colon segmentation. Using episodic training and dual contrastive learning, our Guided-SET-DCL approach achieves remarkable segmentation accuracy, outperforming traditional SOTA FSS methods and CNN-based models. We demonstrated the efficacy of our proposed approach in different training settings that highlighted its robustness and generalization capability. By incorporating sequential episodic training and anatomical guidance, we navigated the complexities of colon segmentation, overcoming challenges such as variable topology and variations in tissue intensity.

References

1. Multi-atlas labeling beyond the cranial vault - workshop and challenge. https://doi.org/10.7303/syn3193805. Accessed 3 Apr 2024
2. Akilandeswari, A., et al.: Automatic detection and segmentation of colorectal cancer with deep residual convolutional neural network. Evid.-Based Complement. Altern. Med. (2022)
3. Alkabbany, I., Ali, A.M., Mohamed, M., Elshazly, S.M., Farag, A.: An AI-based colonic polyp classifier for colorectal cancer screening using low-dose abdominal CT. Sensors **22**(24), 9761 (2022)

4. Arbelle, A., Raviv, T.R.: Microscopy cell segmentation via convolutional LSTM networks. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1008–1012. IEEE (2019)
5. Awate, S.P., Garg, S., Jena, R.: Estimating uncertainty in MRF-based image segmentation: a perfect-MCMC approach. Med. Image Anal. **55**, 181–196 (2019)
6. Bert, A., et al.: An automatic method for colon segmentation in CT colonography. Comput. Med. Imaging Graph. **33**(4), 325–331 (2009)
7. Chen, D., Fahmi, R., Farag, A.A., Falk, R.L., Dryden, G.W.: Accurate and fast 3D colon segmentation in CT colonography. In: ISBI (2009)
8. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. Int. J. Comput. Vision **59**, 167–181 (2004)
9. Gayathri Devi, K., Radhakrishnan, R., et al.: Automatic segmentation of colon in 3D CT images and removal of opacified fluid using cascade feed forward neural network. Comput. Math. Methods Med. **2015** (2015)
10. Guachi, L., Guachi, R., Bini, F., Marinozzi, F., et al.: Automatic colorectal segmentation with convolutional neural network. Comput.-Aided Design Appl. **16**(5), 836–845 (2019)
11. Hanson, M.E., Pickhardt, P.J., Kim, D.H., Pfau, P.R.: Anatomic factors predictive of incomplete colonoscopy based on findings at CT colonography. Am. J. Roentgenol. **189**(4), 774–779 (2007)
12. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: MICCAI multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proceedings of the MICCAI Multi-Atlas Labeling Beyond Cranial Vault-Workshop Challenge, vol. 5, p. 12 (2015)
13. Li, H., Xiong, P., An, J., Wang, L.: Pyramid attention network for semantic segmentation. arXiv preprint arXiv:1805.10180 (2018)
14. Lu, L., Zhang, D., Li, L., Zhao, J.: Fully automated colon segmentation for the computation of complete colon centerline in virtual colonoscopy. IEEE Trans. Biomed. Eng. **59**(4), 996–1004 (2011)
15. Malhotra, P., Gupta, S., Koundal, D., Zaguia, A., Enbeyle, W., et al.: Deep neural networks for medical image segmentation. J. Healthc. Eng. (2022)
16. Mohamad, M., Farag, A., Ali, A.M., Elshazly, S., Farag, A.A., Ghanoum, M.: Enhancing virtual colonoscopy with a new visualization measure. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 294–297. IEEE (2018)
17. Nappi, J.J., Dachman, A.H., MacEneaney, P., Yoshida, H.: 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 (2002)
18. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
19. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervision with superpixels: training few-shot medical image segmentation without annotation. In: ECCV, Part XXIX 16, pp. 762–780. Springer, Cham (2020)
20. Rajamani, K., et al.: Segmentation of colon and removal of opacified fluid for virtual colonoscopy. Pattern Anal. Appl. **21**(1), 205–219 (2018)
21. Ramesh, K., Kumar, G.K., Swapna, K., Datta, D., Rajest, S.S.: A review of medical image segmentation algorithms. EAI Endors. Trans. Pervasive Health Technol. **7**(27), e6–e6 (2021)

22. Ravindran, Z., Das, N.S., et al.: Automatic segmentation of colon using multi-level morphology and thesholding. In: 2021 International Conference on Computer Communication and Informatics (ICCCI), pp. 1–4. IEEE (2021)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: MICCAI 2015, Part III, pp. 234–241. Springer, Cham (2015)
24. Roy, A.G., Navab, N., Wachinger, C.: Recalibrating fully convolutional networks with spatial and channel squeeze and excitation blocks. IEEE Trans. Med. Imaging **38**(2), 540–549 (2018)
25. Roy, A.G., Siddiqui, S., Pölsterl, S., Navab, N., Wachinger, C.: squeeze & excite-guided few-shot segmentation of volumetric images. Med. Image Anal. **59**, 101587 (2020)
26. Sarkar, A., Biswas, M.K., Kartikeyan, B., Kumar, V., Majumder, K.L., Pal, D.: A MRF model-based segmentation approach to classification for multispectral imagery. IEEE Trans. Geosci. Remote Sens. **40**(5), 1102–1113 (2002)
27. Shamir, R.R., Duchin, Y., Kim, J., Sapiro, G., Harel, N.: Continuous dice coefficient: a method for evaluating probabilistic segmentations. arXiv (2019)
28. Tang, H., Liu, X., Sun, S., Yan, X., Xie, X.: Recurrent mask refinement for few-shot medical image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3918–3928 (2021)
29. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: PANet: few-shot image semantic segmentation with prototype alignment. In: proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9197–9206 (2019)
30. Wasserthal, J., et al.: TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. Radiol.: AI (2023)
31. Wu, H., Xiao, F., Liang, C.: Dual contrastive learning with anatomical auxiliary supervision for few-shot medical image segmentation. In: ECCV 2022, pp. 417–434. Springer, Cham (2022)
32. Zhang, C., Lin, G., Liu, F., Yao, R., Shen, C.: CANet: class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5217–5226 (2019)
33. Zhang, H., et al.: ResNest: split-attention networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2736–2746 (2022)
34. Zhang, X., Wei, Y., Yang, Y., Huang, T.S.: SG-one: similarity guidance network for one-shot semantic segmentation. IEEE Trans. Cybern. **50** (2020)

Author Queries

Chapter 5

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding author and email address have been identified as per the information available in the Copyright form.	
AQ2	As Per Springer style, both city and country names must be present in the affiliations. Accordingly, we have inserted the country names in 1 and 4 affiliations. Please check and confirm if the inserted country names are correct. If not, please provide us with the correct country names.	