FSA-Net: Fractal-driven Synergistic Anatomy-aware Network for Segmenting White Line of Toldt in Laparoscopic Images

Kecheng Wu[1], Zhaohu Xing[1], Zerong Cai[2], Feng Gao[2], Wenxue Li[1], and Lei Zhu[1,3] (✉)

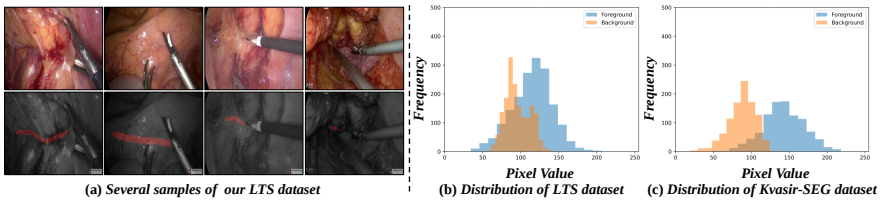
[1] The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
leizhu@ust.hk
[2] The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China
[3] The Hong Kong University of Science and Technology, Hong Kong, Hong Kong

Abstract. Accurate automatic segmentation of the White Line of Toldt (WLT) is crucial for guiding colorectal cancer surgeries and improving patient outcomes. However, the complex anatomical structures and low signal-to-noise ratio involved in relevant regions of WLT pose significant challenges to existing segmentation models. Recent studies highlight fractal dimension as a powerful tool for analyzing the complexity of topological structures, offering an effective approach to representing anatomical features in medical images. Building on its success, we present the first well-annotated laparoscopic WLT segmentation (LTS) dataset and propose FSA-Net, a fractal-driven synergistic anatomy-aware network, specially designed for laparoscopic WLT segmentation. Specifically, FSA-Net consists of two core modules: the local texture-aware convolution (LTC) module and the fractal-guided anatomy-consistent attention (FAA) module. The LTC module adaptively adjusts the convolutional kernel offsets based on fractal dimensions to capture intra-anatomical features, while the FAA module employs a fractal-driven key-value pair filtering strategy to enhance the modeling of correlations across inter-anatomical structures. Extensive experimental results validate the effectiveness of our method. The resources will be available at https://github.com/Bigmouth233/FSA-Net.

Keywords: Fractal Dimension · White Line of Toldt · Laparoscopic Image Segmentation.

1 Introduction

The White Line of Toldt (WLT) is a consistent anatomical structure that marks the junction of rectum and peritoneum, providing an avascular plane for safe mobilization of the mesorectum. Dissection along this plane is essential in procedures such as total mesorectum excision, enabling resection with minimal blood loss while preserving critical structures like the ureter and pelvic nerve [10]. Accurate identification of the WLT is essential during surgeries, as it provides vital



(a) *Several samples of our LTS dataset* (b) *Distribution of LTS dataset* (c) *Distribution of Kvasir-SEG dataset*

Fig. 1. Statistics of our LTS dataset. (a) Several examples of our LTS dataset. (b) Pixel value distribution histogram of our LTS dataset. (c) Pixel value distribution histogram of the Kvasir-SEG dataset. In (b) and (c), claybank represents the background distribution and blue represents the foreground distribution.

guidance for intraoperative decision-making and enhances surgical outcomes. Automatic WLT segmentation algorithms can significantly reduce the burden on surgeons, improve the precision of rectal cancer surgeries and contribute to improving patient prognosis.

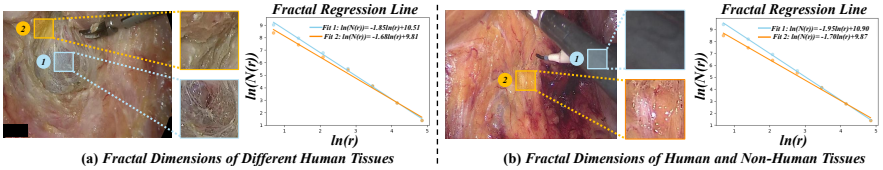
With the advancement of deep learning-driven vision techniques [21,22,24,25], numerous algorithms have been developed for endoscopic image analysis in a variety of clinical scenarios [19,20,27]. However, due to the absence of explicit anatomical constraints, existing models apply a uniform processing strategy to both target and background regions, despite the inconsistent spatial and morphological features of different anatomical structures. Consequently, these models are prone to challenges in laparoscopic WLT segmentation, such as complex anatomical structures and low-contrast targets that are more susceptible to noise.

Recent studies show that fractal geometry offers a robust theory for describing intricate and irregular anatomical features, providing an effective approach to analyze their structural complexity [9]. Inspired by this, we envision that fractal analysis can effectively capture the structural characteristics of various anatomical regions in laparoscopic images, offering a powerful solution for accurate segmentation of the WLT.

To this end, we propose a fractal-driven synergistic anatomy-aware network (FSA-Net), the first to integrate fractal concepts into laparoscopic image segmentation. Specifically, (1) we collect the first high-quality, well-annotated dataset for developing WLT segmentation models. (2) In FSA-Net, we introduce a more efficient algorithm for estimating fractal dimensions. (3) Moreover, we design a local texture-aware convolution (LTC) module to adaptively extract intra-anatomical cues, and a fractal-guided anatomical-consistent attention (FAA) module to capture inter-anatomical context of various anatomical structures. (4) Extensive experimental results on our LTS dataset and polyp segmentation benchmarks demonstrate that our FSA-Net outperforms SOTA methods.

2 Laparoscopic WLT Segmentation (LTS) Dataset

Construction of LTS Dataset. As shown in Fig. 1 (a), we collect 1,715 laparoscopic images with a resolution of $1,920 \times 1,080$ from 145 different colorectal



(a) *Fractal Dimensions of Different Human Tissues* (b) *Fractal Dimensions of Human and Non-Human Tissues*

Fig. 2. Fractal analysis of different regions in laparoscopic images. According to the box-counting theory, the dimension can be estimated by performing a linear fit of $\ln(N(r))$ against $\ln(r)$, where $N(r)$ is the quantity of boxes and r is the diameter of each box. (a) Comparison of estimation results across different anatomical structures. (b) Comparison of estimation results between human and non-human structures.

cancer patients. Experienced surgeons performed pixel-level annotations to ensure precise delineation of the WLT. The dataset is divided into a training set containing 1,372 images and a test set consisting of 343 images.

Dataset Analysis. The pixel value distributions of our LTS dataset and the Kvasir-SEG dataset [6] are shown in Fig. 1 (b) and (c). Notably, the pixel value distributions of our LTS dataset display a significant overlap, indicating a lower contrast between the WLT and the surrounding background.

3 Method

3.1 Preliminary and Motivation

The Hausdorff dimension is a key concept in fractal geometry, providing an effective measure of texture complexity [14], and is defined as:

$$\begin{aligned} \mathcal{H}^z(S) &= \lim_{\delta \to 0} \inf \left\{ \sum_i (\operatorname{diam} U_i)^z \mid S \subseteq \bigcup_i U_i, \operatorname{diam} U_i < \delta \right\}, \\ \dim_H(S) &= \inf\{z \geq 0 \mid \mathcal{H}^z(S) = 0\}, \end{aligned} \tag{1}$$

where the collection $\{U_i\}$ represents an arbitrary countable or finite cover of S, and diam U_i denotes the diameter of U_i. The box-counting method is the most commonly used technique for estimating fractal dimensions in practice [8]. As illustrated in Fig. 2, a preliminary fractal analysis of surgical laparoscopic images using the box-counting method reveals obvious variations in fractal dimensions across different anatomical structures. Motivated by above findings, we propose integrating fractal dimension features into our model to better capture and exploit the distinct characteristics of various anatomical structures.

3.2 Fractal Dimension Estimation

While the box-counting method [8] is effective for estimating fractal dimensions across different structures, the feature maps extracted by deep learning model

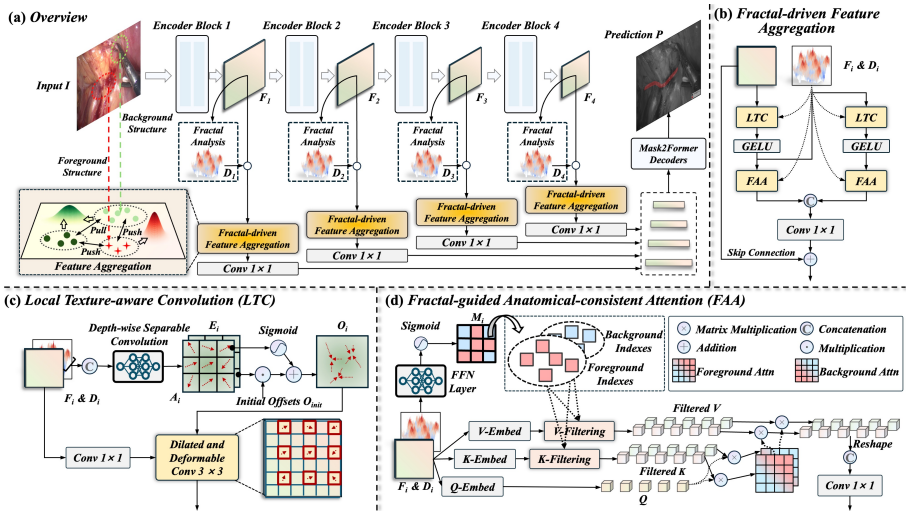


Fig. 3. Architecture of the proposed FSA-Net. (a) The overview of our FSA-Net. (b) The fractal-driven synergistic feature aggregation pipeline. It contains two core modules, the LTC module and the FAA module. (c) The details of the LTC module. The LTC module use the dilated and deformable convolution for intra-anatomical feature aggregation. (d) The details of FAA module. The FAA module adopt the proposed key-value filtering attention to effectively extract inter-anatomical context.

backbones often have high embedding dimensions, resulting in significant computational complexity. In contrast, measure-based fractal dimension estimation methods (e.g., Rényi generalized dimension [17]) are generally more robust when dealing with high-dimensional data, effectively mitigating the curse of dimensionality associated with direct geometric covering.

To this end, we propose a novel algorithm based on the q-th order correlation integral [15] for estimating fractal dimensions at the pixel level of an image. The algorithm details are shown in Algo. 1. Specifically, $\mathcal{B}(F_i[h,w],\epsilon)$ represents the $\epsilon \times \epsilon$ neighborhood centered at the pixel $F_i[h,w]$, $\mathbb{E}[\cdot]$ denotes the mathematical expectation operator, q is the scaling exponent which is set to 2 for computational convenience, and $\mathrm{Dis}(\cdot)$ is the distance measurement of two F_i pixels:

$$\mathrm{Dis}(x,y) = \frac{1}{\pi}\arccos\left(\frac{\sum_{c=1}^{C_i} x_c \cdot y_c}{\sqrt{\sum_{c=1}^{C_i}(x_c)^2}\sqrt{\sum_{c=1}^{C_i}(y_c)^2}}\right) + eps \text{ and } x, y \in F_i, \quad (2)$$

where the subscript c denotes the c-th channel of the input pixel, eps is a small constant for preventing division by zero. Given the computational complexity of limit-based calculations, we approximate the fractal dimension by adopting the slope of the linear function fitted between $\ln(\mathcal{C}_q^\epsilon)$ and $\ln(\epsilon)$. We implement Algo. 1 in PyTorch, optimizing the process using matrix operations.

Algorithm 1. Fractal Dimension Estimation

Require: Feature map F_i with size $H_i \times W_i$; Neighborhood scale parameter ϵ
Padding Size $p \longleftarrow \lfloor \epsilon/2 \rfloor$
$F_i^{pad} \longleftarrow$ Reflect Padding(F_i, p)
Fractal Dimension Map $D_i \longleftarrow$ Zero Tensor of Shape $1 \times H_i \times W_i$
For $h \in$ range$(0, H_i)$ do:
For $w \in$ range$(0, W_i)$ do:
$R_i^{(h,w)} \longleftarrow \mathcal{B}(F_i[h, w], \epsilon)$ and $\mathcal{B}(F_i[h, w], \epsilon) \subseteq F_i^{pad}$
$\mathcal{C}_q^{\epsilon} \leftarrow \int_{x \in R_i^{(h,w)}} \left(\frac{\epsilon}{\mathbb{E}[\mathrm{Dis}(x,y) | y \in R_i^{(h,w)}]} \right)^{q-1} d\mu(x)$ and $d_q^{(h,w)} \leftarrow \lim_{\epsilon \to 0} \frac{\ln(\mathcal{C}_q^{\epsilon})}{(q-1)\ln(\epsilon)}$
$D_i[h, w] \longleftarrow d_q^{(h,w)}$
return D_i

3.3 Overall Architecture

As shown in Fig. 3 (a), given an image $I \in \mathbb{R}^{3 \times H \times W}$, we adopt Swin-$S$ [12] as the backbone to extract multi-scale features $\{F_i\}_{i=1}^4 \in \mathbb{R}^{C_i \times H_i \times W_i}$. The fractal dimensions of these features are then estimated. Guided by these fractal dimensions, a feature aggregation pipeline is applied to the backbone features. Finally, the updated features are passed through the decoder [2], and the mask $P \in \mathbb{R}^{1 \times H \times W}$ is generated by the decoder head. For the loss function, we follow the protocol described in [5].

3.4 Fractal-driven Synergistic Feature Aggregation

Fractal features effectively represent anatomical features and offer precise analysis of the microstructure of different structures. Inspired by this, we leverage fractal information as guidance to adaptively capture both intra- and inter-anatomical context in our model. The detailed pipeline for fractal-driven synergistic feature aggregation is shown in Fig. 3 (b). For simplicity, we use the same symbol F_i to represent the backbone features within the pipeline.

Local Texture-aware Convolution (LTC) Module. The LTC module is designed for the adaptive aggregation of intra-anatomical features of different anatomical structures. Unlike standard DCNs, we constrain the offsets of convolution kernels using anatomical knowledge embedded in fractal features, ensuring that the model avoids learning unreasonable deformations. As shown in Fig. 3 (c), we concatenate F_i and D_i as input to predict the offsets of the DCN kernel. The offset generator consists of two 3×3 depth-wise separable convolution blocks to predict the dilatation coefficients $E_i \in \mathbb{R}^{k^2 \times H_i \times W_i}$ and the offset directions $A_i \in \mathbb{R}^{2k^2 \times H_i \times W_i}$, where k is the kernel size of the DCN layer. Subsequently, E_i and A_i are used to compute the final offset $O_i \in \mathbb{R}^{2k^2 \times H_i \times W_i}$. The dilated and deformable convolution then uses F_i and O_i to calculate the output features.

Fractal-guided Anatomical-consistent Attention (FAA) Module. We propose a novel FAA module to effectively capture inter-anatomical reciprocal action. The core of the FAA module is a fractal-guided key-value pair filtering

strategy, designed to enhance the model's ability to capture inter-anatomical information in a divide-and-conquer manner. Specifically, we first employ a feed-forward neural network (FFN) to transform D_i into a normalized distribution map $M_i \in \mathbb{R}^{2\times H_i \times W_i}$. Then, we extract the indexes of foreground regions (e.g., WLT-related regions) and background regions (e.g., surgical instruments or distal tissues in surgical field) based on M_i, respectively:

$$\mathcal{I}_i^f = \{(h,w)|M_i[h,w] \geq 0.5\} \; and \; \mathcal{I}_i^b = \{(h,w)|M_i[h,w] < 0.5\}, \tag{3}$$

where $\mathcal{I}_i^f$ represents the set of foreground indexes and $\mathcal{I}_i^b$ represents the set of background indexes. Then, the input feature F_i is flattened to obtain $F_i^{fla} \in \mathbb{R}^{H_i W_i \times C_i}$ for the embeddings of Q, K and V. Subsequently, for each head $t \in 1, 2, ..., n$, we define the projection matrices $\mathcal{W}_Q^{t|i}, \mathcal{W}_K^{t|i}, \mathcal{W}_V^{t|i} \in \mathbb{R}^{C_i \times \frac{C_i}{n}}$. Taking the foreground as an example, the core procedure of FAA can be expressed as:

$$F_i^f = \mathrm{Cat}\left(\left\{\mathcal{S}\left(\frac{F_i^{fla}\mathcal{W}_Q^{t|i}(F_i^{fla}\lfloor\mathcal{I}_i^f\rceil\mathcal{W}_K^{t|i})^\top}{\sqrt{C_i/n}}\right)F_i^{fla}\lfloor\mathcal{I}_i^f\rceil\mathcal{W}_V^{t|i}\right\}_{t=1}^{n}\right), \tag{4}$$

where $\mathrm{Cat}(\cdot)$ is the concatenation operation, $\mathcal{S}$ represents the Softmax function for normalization, and $\lfloor\cdot\rceil$ is the key-value pairs filtering process. The computation for the background feature F_i^b follows a similar procedure as for F_i^f. Finally, F_i^f and F_i^b are concatenated, and a Conv 1×1 is applied for feature fusion.

4 Experiments

4.1 Datasets and Implementation Details

The LTS Dataset. Our LTS dataset consists of 1,715 laparoscopic images from colorectal surgeries, each with a resolution of 1,920 × 1,080. The dataset is randomly divided into a training set and a testing set at an 8:2 ratio, resulting in 1,372 images for training and 343 images for testing.

Polyp Segmentation Benchmarks. We also conduct extensive experiments on the polyp segmentation datasets outlined in [7], which serve as widely used benchmarks for all types of polyp segmentation models [5,7]. The training set consists of 6,925 images. We adopt the CVC-300-TV and CVC-612-V datasets as two independent test sets to evaluate all methods, ensuring a comprehensive assessment of segmentation performance across diverse polyp samples.

Implementation Details. The proposed FSA-Net is implemented in PyTorch 2.1.1 with CUDA 11.8. The input image is resized to 384×384, and a batch size of 4 is used, with training conducted for 100 epochs on each dataset. The AdamW optimizer is employed, along with a cosine annealing learning rate scheduler. The initial learning rate is set to 1×10^{-4}, decaying to 1×10^{-6}. Data augmentation techniques, including vertical and horizontal flipping, rotation, gamma correction, and elastic deformation, are applied to enhance model generalization. We adopt Dice coefficient, Intersection over Union (IoU), weighted F-measure (F_β^w) [13], mean absolute error (MAE), S-measure (S_α) [3], and E-measure (E_ϕ) [4] as metrics to assess our segmentation results.

Table 1. Quantitative comparison with SOTA methods on our LTS dataset. The best scores are highlighted in bold and the suboptimal scores are underlined.

Methods	Image Type	Venue	Years	Metrics					
				Dice ↑	IoU ↑	F_β^w ↑	S_α ↑	E_ϕ ↑	MAE ↓
UNet	General	MICCAI	2015	0.413	0.305	0.363	0.575	0.807	0.026
UNet++	General	TMI	2019	0.452	0.335	0.401	0.593	0.830	0.025
ACSNet	Endoscopic	MICCAI	2020	0.476	0.353	0.424	0.606	0.836	0.024
PraNet	Endoscopic	MICCAI	2020	0.511	0.387	0.482	0.629	0.884	0.020
TransUNet	General	arXiv	2021	0.517	0.388	0.460	0.623	0.822	0.024
SCR-Net	Endoscopic	AAAI	2021	0.443	0.331	0.399	0.595	0.834	0.022
LDNet	Endoscopic	MICCAI	2022	0.516	0.390	0.478	0.628	0.885	0.020
CASCADE	General	WACV	2023	0.528	0.400	<u>0.494</u>	0.633	<u>0.887</u>	<u>0.018</u>
Swin-UMamba	General	MICCAI	2024	<u>0.533</u>	<u>0.405</u>	0.492	<u>0.634</u>	0.877	0.020
I^2Net	General	MICCAI	2024	0.489	0.365	0.443	0.614	0.852	0.022
FSA-Net	Endoscopic	(Ours)	2025	**0.557**	**0.428**	**0.507**	**0.650**	**0.902**	**0.016**

Table 2. Quantitative comparison with SOTA methods on polyp segmentation benchmarks. The best scores are highlighted in bold.

Datasets	Metrics	UNet MICCAI 2015	UNet++ TMI 2019	ACSNet MICCAI 2020	PraNet MICCAI 2020	TransUNet arXiv 2021	LDNet MICCAI 2022	CASCADE WACV 2023	Swin-UMamba MICCAI 2024	FSA-Net (Ours) 2025
CVC-300-TV	Dice↑	0.639	0.649	0.738	0.739	0.824	0.835	0.844	0.850	**0.878**
	IoU↑	0.525	0.539	0.632	0.645	0.735	0.741	0.751	0.759	**0.790**
	S_α ↑	0.793	0.796	0.837	0.833	0.872	0.898	0.902	0.910	**0.920**
	E_ϕ ↑	0.826	0.831	0.871	0.852	0.895	0.910	0.925	0.931	**0.955**
	MAE ↓	0.027	0.024	0.016	0.016	0.016	0.015	0.014	**0.012**	0.013
CVC-612-V	Dice↑	0.725	0.684	0.804	0.869	0.861	0.870	0.878	0.875	**0.888**
	IoU↑	0.610	0.570	0.929	0.799	0.780	0.799	0.811	0.803	**0.817**
	S_α ↑	0.826	0.805	0.847	0.915	0.893	0.918	0.920	0.915	**0.930**
	E_ϕ ↑	0.855	0.830	0.887	0.936	0.921	0.941	0.942	0.946	**0.962**
	MAE ↓	0.023	0.025	0.054	0.013	0.015	0.013	0.013	0.014	**0.012**

4.2 Comparisons with SOTA methods

WLT Segmentation. We compare our FSA-Net against ten SOTA segmentation methods, including UNet [18], UNet++ [30], ACSNet [29], PraNet [5], TransUNet [1], SCR-Net [23], LDNet [28], TransCASCADE [16], Swin-UMamba [11], and I^2Net [26]. As shown in Tab. 1, among all the comparison methods, as a Mamba-based method, Swin-UMamba achieves the highest Dice and IoU scores (0.533 and 0.405). Compared to Swin-UMamba, our FSA-Net improves the Dice score and the IoU score by 4.50% and 5.68%, respectively. Moreover, our approach also achieves the best performance across the other four scores by a considerable margin. The comparison results demonstrate the advancement of our FSA-Net among general and colonoscopic image segmentation models.

Polyp Segmentation. we further compare FSA-Net against SOTA image-based methods on public benchmarks for polyp segmentation. Tab. 2 presents the quantitative segmentation results on the CVC-300-TV and CVC-612-V. Our FSA-Net achieves superior performance across nearly all metrics compared to other methods on both datasets and demonstrates strong generalization ability.

Visual Comparisons. Fig. 4 visually compares the WLT and polyp segmentation results of FSA-Net with those of six other methods. The results demonstrate

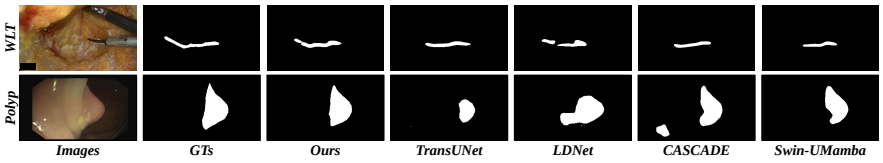


Fig. 4. Qualitative comparison results on two samples from our LTS dataset and polyp segmentation benchmarks with some of the SOTA methods.

Table 3. Ablation study for the two different core modules (LTC and FAA) on two datasets.

Models	Core Modules		LTS		CVC-612-V	
	LTC	FAA	Dice ↑	IoU ↑	Dice ↑	IoU ↑
Basic	✗	✗	0.526	0.394	0.871	0.799
B_1	✓	✗	0.539	0.405	0.875	0.802
B_2	✗	✓	0.547	0.412	0.880	0.811
FSA-Net	✓	✓	**0.557**	**0.428**	**0.888**	**0.817**

Table 4. Ablation study for different feature extraction strategies on LTS dataset. DCN denotes the deformable convolution and MSA denotes the vanilla multi-head self-attention.

Models	Core Modules	Is Fractal-driven	Metrics	
			Dice ↑	IoU ↑
B_3	DCN	No	0.533	0.399
B_4	MSA	No	0.528	0.394
B_5	DCN & MSA	No	0.536	0.403
FSA-Net	LTC & FAA	Yes	**0.557**	**0.428**

that our method achieves more accurate recognition of the WLT region and better detects the boundaries of polyps with irregular textures and shapes.

4.3 Ablation Study

Effectiveness of LTC and FAA Modules. We conduct ablation studies by systematically removing each component from FSA-Net and evaluating its impact on both the LTS dataset and the polyp segmentation benchmark. As shown in Tab. 3, "*Basic*" represents the baseline of our method, consisting solely of the encoder-decoder architecture. In B_1, we incorporate the LTC modules, while in B_2, we introduce the FAA modules into the "*Basic*". Both B_1 and B_2 demonstrate significant improvements in Dice and IoU scores on both datasets.
Effectiveness of Fractal Guidance. We further construct three additional baseline networks: B_3, B_4, and B_5. As shown in Tab. 4, in B_3, we replace the LTC module in FSA-Net with a vanilla deformable convolution module, while in B_4, we replace the FAA module with a self-attention module. In B_5, we replace both modules with their respective standard counterparts. Experimental results on the LTS dataset reveal a notable decline in segmentation performance following these replacements. These ablation studies emphasize the effectiveness of integrating fractal information, demonstrating that its incorporation significantly enhances the segmentation performance of segmentation models.

5 Conclusion

In this paper, we propose FSA-Net, the first framework for segmenting the White Line of Toldt (WLT) in surgical laparoscopic images. First, we introduce a novel fractal dimension estimation algorithm for accurately computing fractal dimensions. Next, we propose the local texture-aware convolution (LTC) module and

the fractal-guided anatomical-consistent attention (FAA) module to synergistically extract intra- and inter-anatomical features. Additionally, we construct the first laparoscopic WLT segmentation (LTS) dataset to support related research. Extensive experiments on both the LTS and polyp segmentation benchmarks demonstrate the effectiveness of our approach.

Acknowledgments. This research is supported by the Guangdong Science and Technology Department (No. 2024ZDZX2004), the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0501600), the National Natural Science Foundation of China (No.8227242,FG), the Guangzhou Basic and Applied Basic Research Fund (No.2024A04J9983,FG), the Guang-Dong Basic and Applied Basic Research Foundation (2024A1515220041), and the Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things(No.2023B1212010007).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
2. Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R.: Masked-attention mask transformer for universal image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1290–1299 (2022)
3. Fan, D.P., Cheng, M.M., Liu, Y., Li, T., Borji, A.: Structure-measure: A new way to evaluate foreground maps. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4548–4557 (2017)
4. Fan, D.P., Gong, C., Cao, Y., Ren, B., Cheng, M.M., Borji, A.: Enhanced-alignment measure for binary foreground map evaluation. arXiv preprint arXiv:1805.10421 (2018)
5. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranet: Parallel reverse attention network for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 263–273. Springer (2020)
6. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., De Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: MultiMedia Modeling: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II 26. pp. 451–462. Springer (2020)
7. Ji, G.P., Chou, Y.C., Fan, D.P., Chen, G., Fu, H., Jha, D., Shao, L.: Progressively normalized self-attention network for video polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 142–152. Springer (2021)

8. Konatar, I., Popovic, T., Popovic, N.: Box-counting method in python for fractal analysis of biomedical images. In: 2020 24th International Conference on Information Technology. pp. 1–4. IEEE (2020)
9. Lee, W.L., Hsieh, K.S.: A robust algorithm for the fractal dimension of images and its applications to the classification of natural images and ultrasonic liver images. Signal Processing **90**(6), 1894–1904 (2010)
10. Li, K., He, X., Zheng, Y.: An optimal surgical plane for laparoscopic functional total mesorectal excision in rectal cancer. Journal of Gastrointestinal Surgery **25**(10), 2726–2727 (2021)
11. Liu, J., Yang, H., Zhou, H.Y., Xi, Y., Yu, L., Li, C., Liang, Y., Shi, G., Yu, Y., Zhang, S., et al.: Swin-umamba: Mamba-based unet with imagenet-based pretraining. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 615–625. Springer (2024)
12. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022 (2021)
13. Margolin, R., Zelnik-Manor, L., Tal, A.: How to evaluate foreground maps? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2014)
14. Pentland, A.P.: Fractal-based description of natural scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence (6), 661–674 (1984)
15. Procacia, I., et al.: Measuring the strangeness of strange attractors. Physica. D **9**(1-2), 189–208 (1983)
16. Rahman, M.M., Marculescu, R.: Medical image segmentation via cascaded attention decoding. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6222–6231 (2023)
17. Rényi, A.: On the dimension and entropy of probability distributions. Acta Mathematica Academiae Scientiarum Hungarica **10**, 193–215 (1959)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
19. Wang, H., Jin, Y., Zhu, L.: Dynamic interactive relation capturing via scene graph learning for robotic surgical report generation. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 2702–2709. IEEE (2023)
20. Wang, H., Yang, G., Zhang, S., Qin, J., Guo, Y., Xu, B., Jin, Y., Zhu, L.: Video-instrument synergistic network for referring video instrument segmentation in robotic surgery. IEEE Transactions on Medical Imaging (2024)
21. Wu, H., Yang, Y., Aviles-Rivero, A.I., Ren, J., Chen, S., Chen, H., Zhu, L.: Semi-supervised video desnowing network via temporal decoupling experts and distribution-driven contrastive regularization. In: European Conference on Computer Vision. pp. 70–89. Springer (2024)
22. Wu, H., Yang, Y., Xu, H., Wang, W., Zhou, J., Zhu, L.: Rainmamba: Enhanced locality learning with state space models for video deraining. arXiv preprint arXiv:2407.21773 (2024)
23. Wu, H., Zhong, J., Wang, W., Wen, Z., Qin, J.: Precise yet efficient semantic calibration and refinement in convnets for real-time polyp segmentation from colonoscopy videos. In: Association for the Advancement of Artificial Intelligence. vol. 35, pp. 2916–2924 (2021)

24. Xing, Z., Wan, L., Fu, H., Yang, G., Yang, Y., Yu, L., Lei, B., Zhu, L.: Diff-unet: A diffusion embedded network for robust 3d medical image segmentation. Medical Image Analysis p. 103654 (2025)
25. Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 578–588. Springer (2024)
26. Yu, J., Duan, F., Chen, L.: I 2 net: Exploiting misaligned contexts orthogonally with implicit-parameterized implicit functions for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 328–338. Springer (2024)
27. Yue, G., Li, Y., Jiang, W., Zhou, W., Zhou, T.: Boundary refinement network for colorectal polyp segmentation in colonoscopy images. IEEE Signal Processing Letters (2024)
28. Zhang, R., Lai, P., Wan, X., Fan, D.J., Gao, F., Wu, X.J., Li, G.: Lesion-aware dynamic kernel for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 99–109. Springer (2022)
29. Zhang, R., Li, G., Li, Z., Cui, S., Qian, D., Yu, Y.: Adaptive context selection for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 253–262. Springer (2020)
30. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging **39**(6), 1856–1867 (2019)