LACFormer: Toward accurate and efficient polyp segmentation

Authors: Quan Nguyen, Mai Nguyen, Thanh Tung Nguyen, Huy Quang Trinh, Linh Bao Doan, Toan Van Pham

R&D Lab, Sun* Inc

Abstract

Polyp segmentation is an essential task in medical image analysis for early detection of colorectal cancer. Deep learning models, particularly encoder-decoder architectures, have been successful in polyp segmentation. However, these models often struggle to capture long-range dependencies and exhibit limited performance on small polyps. In this paper, we propose LACFormer, a novel hierarchical Transformer-CNN model incorporating the Laplacian pyramid for polyp segmentation. The proposed model combines the strengths of Transformers and CNNs along with Laplacian images to overcome the limitations of previous models. Specifically, the hierarchical Transformer backbone captures long-range dependencies and hierarchically processes the features to generate multi-scale representations. These representations are then fused with a novel CNN decoder, which enhances feature representations and refines the segmentation masks. Besides, many novel modules for effective polyp segmentation are also proposed. We evaluated our model on five popular benchmark datasets for polyp segmentation, including Kvasir, CVC-Clinic DB, CVC-ColonDB, CVC-T, and ETIS-Larib. Experimental results show that LACFormer outperforms state-of-the-art models, achieving a Dice similarity coefficient (DSC) of 0.927 and a mean intersection-over-union (mIoU) of 0.878 on CVC-ClinicDB, a DSC of 0.831 and mIoU of 0.753 on CVC-ColonDB and a DSC of 0.824 and mIoU of 0.753 on ETIS-Larib.

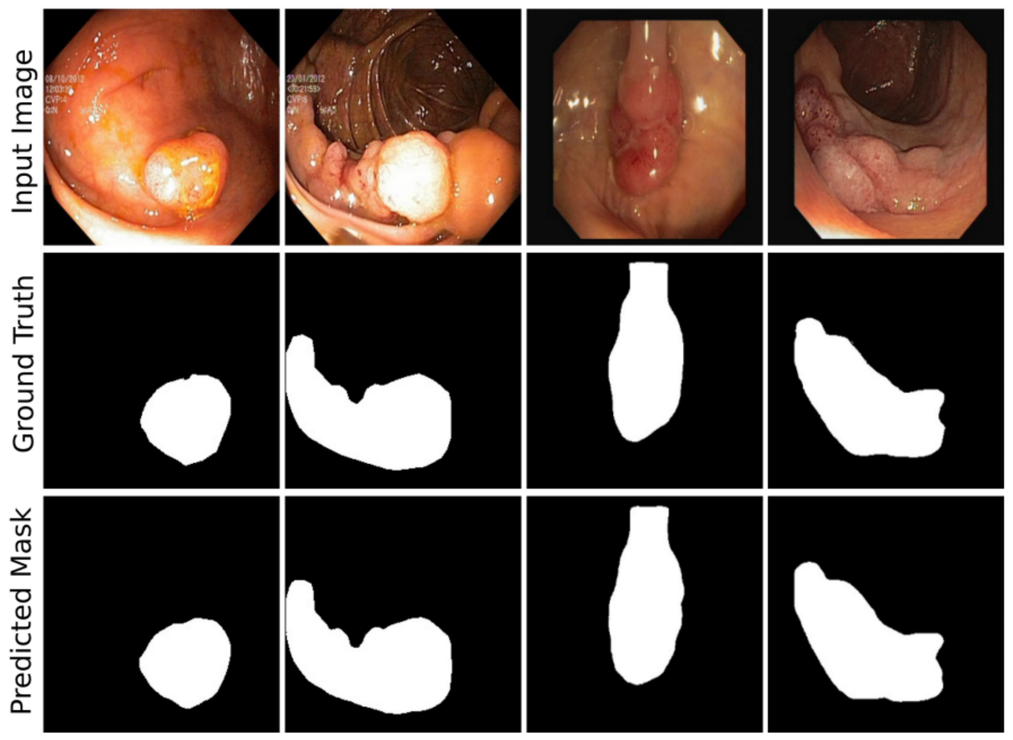


Figure 1. Qualitative results of the LACFormer

Contributions

⇒ We propose a **L**aplacian **A**trous **C**ascaded Transformer(LACFormer) model for polyp segmentation task that is capable of effectively capturing polyps of various sizes.

⇒ We design a novel decoder head suitable for hierarchical encoder transformer architecture with newly developed modules: Atrous Sequential Aggregation Module (ASAM), and Scale Guidance.

⇒ Our proposed LACFormer improves the SOTA performance on CVC-ClinicDB, CVC-ColonDB, and ETIS-LaribLarib by 1%, 2% and 2.9% respectively.

Method

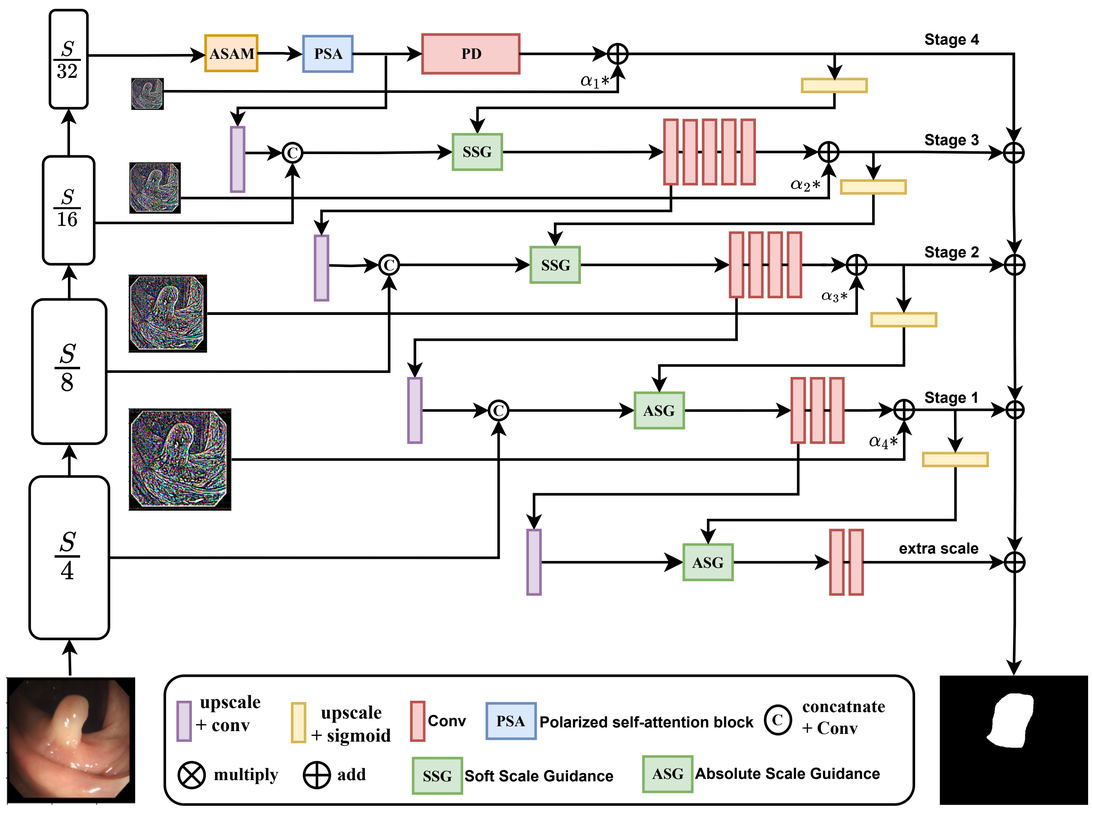
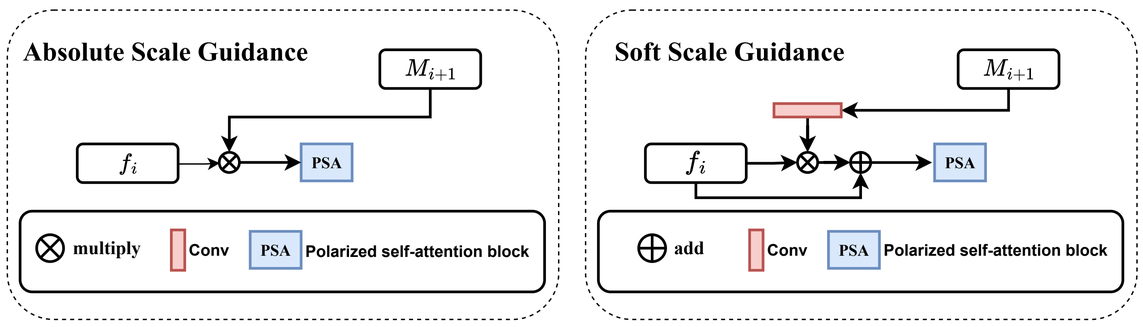
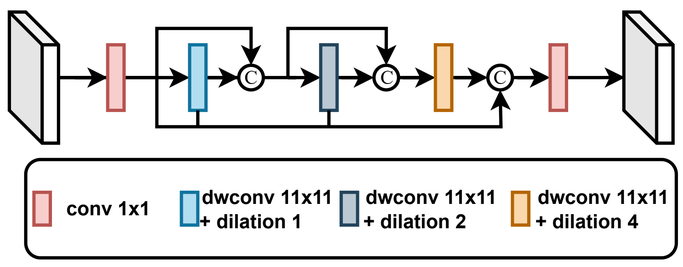


Figure 2. The architecture of proposed LACFormer. "Conv" denotes pre-activation based convolution block: BatchNorm + ReLU + Convolution. "Upscale + conv" is pre-activation based upsampling convolution block: BatchNorm + ReLU + Upsampling + Convolution. "Upscale + sigmoid" is normally an upscale operation then sigmoid.



(a) Scale Guidance



(b) Atrous Sequential Aggregation Module

Figure 3. Sub-module of LACFormer model

Dataset

We conduct experiments on five polyp segmentation datasets: Kvasir , CVC-ClinicDB, CVC-ColonDB, CVC-T and ETIS-Larib Polyp DB. We randomly extract 1450 images both from Kvasir and CVC-ClinicDB to construct a training dataset. Then we perform evaluation on the rest of Kvasir and CVC-ClinicDB. We also evaluate on CVC-ColonDB, CVC-T, and ETIS-Larib which relatively contain 380 images, 60 images and 196 images

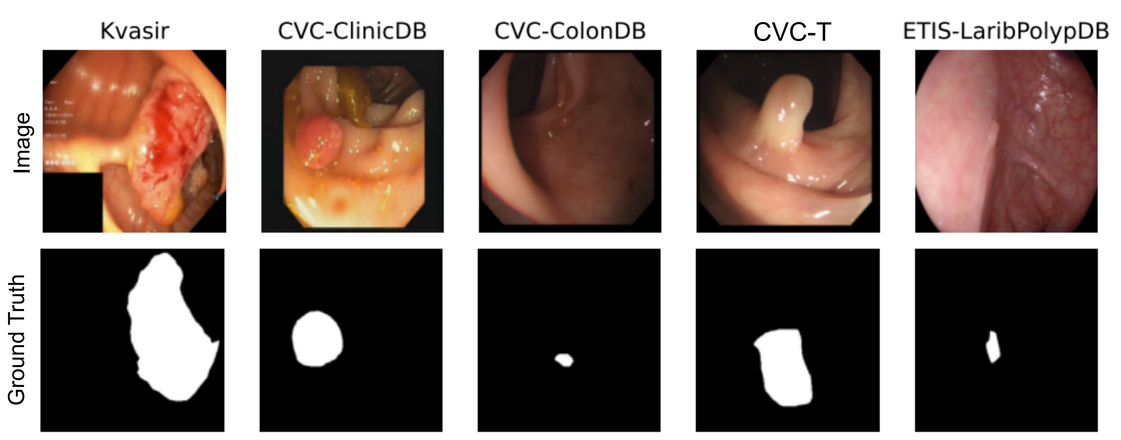


Figure 4. Datasets for experimenting.

Table 1. Statistics of five polyp segmentation datasets.

Dataset	Average Resolution	Train Samples	Test Samples
Kvasir	618 x 539	838	100
CVC-ClinicDB	384 x 288	612	62
CVC-ColonDB	574 x 500	0	380
CVC-T	574 x 500	0	60
ETIS-LaribPolypDB	1225 x 966	0	196

Experimental Results

We compare our results with existing approaches on 5 benchmark datasets. Table 2 shows the results of SOTA methods.

Table 2. Comparison with other approaches on 5 benchmark datasets

Methods	Kvasir		ClinicDB		ColonDB		CVC-T		ETIS	
	mDice	mIou	mDice	mIou	mDice	mIou	mDice	mIou	mDice	mIou
PraNet [12]	0.898	0.840	0.899	0.849	0.709	0.640	0.871	0.797	0.628	0.567
Polyp-PVT [9]	0.917	0.864	0.937	0.889	0.808	0.727	0.900	0.833	0.787	0.706
SANet [36]	0.904	0.847	0.916	0.859	0.753	0.670	0.888	0.815	0.750	0.654
MSNet [42]	0.907	0.862	0.921	0.879	0.755	0.678	0.869	0.807	0.719	0.664
TransFuse-L* [40]	0.920	0.870	**0.942**	**0.897**	0.781	0.706	0.894	0.826	0.737	0.663
SSFormer-L [33]	0.917	0.864	0.906	0.855	0.802	0.721	0.895	0.827	0.796	0.720
ColonFormer-L [11]	0.924	0.876	0.932	0.884	0.811	0.733	**0.906**	**0.842**	0.801	0.722
LACFormer-L (Ours)	**0.927**	**0.878**	0.932	0.885	**0.831**	**0.753**	0.892	0.825	**0.824**	**0.753**

Conclusion

In this work, we propose a novel deep neural network architecture called LACFormer for colon polyp segmentation. The proposed approach holds great potential in applications of laplacian image for medical image analysis. Together with Laplacian Pyramid, Atrous Sequential Aggregation Module and polarize self-attention also play an important role in searching and refining potential polyp regions. The experimental results on the five public datasets demonstrate the significant performance of our model compared to state-of-the-art methods.

nguyen.van.quan@sun-asterisk.com