Strong-Guided Pixel Supervised Contrast for Polyp Segmentation

Ledan Tan
Guangdong University of Technology
Guangzhou, China
m13686959306@163.com

Hai-Lin Liu
Guangdong University of Technology
Guangzhou, China
hlliu@gdut.edu.cn

Lei Chen
Guangdong University of Technology
Guangzhou, China
chenlei3@gdut.edu.cn

Yuping Wang
Xidian University
Shaanxi, China
ywang@xidian.ed.cn

***Abstract*—The current polyp segmentation methods mainly use the saliency map to obtain the uncertain region, foreground region, and background region of the polyp image, and then they learn the semantic information from each other, to enhance the edge segmentation ability of the network. However, there is great instability in the quality of the saliency map and the error information brought by low-quality saliency maps will interfere with the segmentation ability of the network. To this end, this paper proposes a strong-guided, pixel-wise, supervised contrastive learning method (SGP-SCL), which enhance the model to identify the polyp boundary by strengthening foreground and background guidance for polyp boundary. Specifically, the SGPS-CL method fully utilizes the ground truth label to obtain high-confidence and representative samples to guide the learning of boundary regions with low confidence, thus reducing the impact of the instability of the preliminary prediction probability map quality on the network performance. Experiments are conducted on CVC-300, CVCClinicDB, Kvasir, CVC-ColonDB, and ETIS polyp segmentation datasets, and the proposed method achieves competitive results.**

Keywords—polyp segmentation, contrast learning, strong-Guided

I. INTRODUCTION

Colorectal polyps, as significant precancerous lesions, hold the key to averting the onset of colorectal cancer through timely detection and removal [1]. Historically, manual screening via colonoscopy has proven effective in identifying precursor lesions and early stages of colorectal cancer [2]. Nevertheless, the efficacy of colonoscopy screening is heavily reliant on the clinician's clinical surgical experience and professional knowledge reservoir. Prolonged exposure to this work can lead to subjective judgment and emotional influences, potentially resulting in instances of missed diagnoses and misdiagnoses of polyps. Consequently, provide automated, accurate, and computer-aided detection techniques with dependable predictive capabilities becomes paramount.

In recent years, deep learning-based semantic segmentation have proven highly effective for diagnosing colorectal lesions, owing to their emphasis on pixel-wise classification and dense prediction [3]. For example, UNet [4] uses the U-shaped architecture and the jump connection to join the semantic information and the deeper features, so that the details are recovered more refined. And UNet++ [5] based on UNet variants are also proposed. Through semantic segmentation, clinicians and researchers can swiftly and precisely identify the location and extent of colorectal lesions. Nevertheless, polyp segmentation presents two formidable challenges [6]. On one hand, colorectal polyps exhibit diverse sizes, shapes, and appearances. On the other hand, the similarities in color, shape, and texture between polyp foreground and background result in indistinct boundaries. To address the first challenge, some methods incorporate multi-scale feature fusion to enhance network robustness. SFANet [7], for instance, introduces adaptive scale context module and semantic global context module to extract multi-scale context features, thereby bolstering feature fusion between high-level and low-level features. Polypseg [8] adopts a selective kernel module, enabling the adaptive selection of kernels of various sizes to extract multiple receptive field features. While these approaches allow the polyp segmentation network to focus on multi-scale information and improve detailed information learning, they do not entirely resolve the issue of fuzzy boundaries. Researchers have also explored methods to address the challenge of fuzzy boundaries. For instance, in [9], a dual-branch structure is employed to formalize boundary-sensitive loss by inferring the target boundary through the region branch, and predicting the target contour through the boundary branch. Another approach, proposed in [10], introduces a novel boundary constraint network (BCNet) that achieves more precise polyp segmentation through joint supervision of polyp regions and boundaries. However, it's important to note that this method, which incorporates additional boundary segmentation information. Recently, some studies have incorporated the concept of saliency detection to identify objects in polyp segmentation images, and subsequently enhance target boundaries in a bottom-up fashion. For instance, the Parallel Reverse Attention Network (PraNet) [6] was introduced, which aggregates features into the decoder to generate a guided saliency map. The reverse attention module establishes relationships between boundaries and regions, extracting boundary clues and thereby improving segmentation accuracy. Another approach, the Uncertainty Augmented Context Attention Network (UACANet) [11] partitions the foreground, background, and uncertain regions using the saliency map. Following mapping and aggregation, the network enhances the context representation ability of the uncertain region through a novel self-attention mechanism. However, relying solely on the saliency map to devide image foreground,

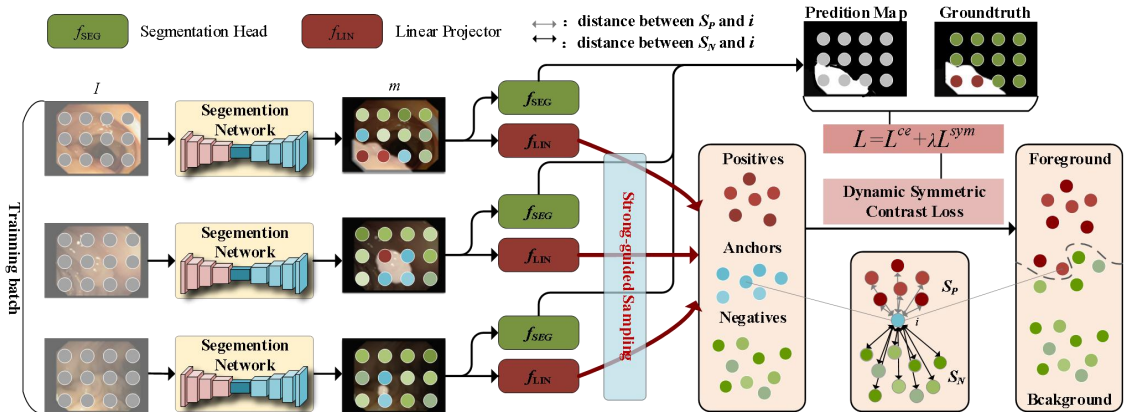

Figure.1. The framework of the proposed method.

background, and uncertain regions for contextual semantic learning is deemed unreliable. The inaccuracies in the saliency map can result in insufficient guided learning ability of the network, thereby impacting overall performance. Consequently, mitigating the negative impact caused by saliency map misguidance, and effectively directing the uncertain region become pivotal in addressing the challenge of fuzzy boundaries in colorectal polyp segmentation.

Fortunately, some researchers have introduced Cross-Image Pixel-wise Supervised Contrastive Learninng (CIP-SCL) [12], which does not require saliency maps for segmentation. CIP-SCL utilizes the ground truth label of each pixel to determine positive and negative samples, learning feature similarity in the embedding space to enhance global feature constraints, and improve network segmentation performance. However, CIP-SCL also poses limitations in the context of polyp segmentation. The sampling strategy involves hard samples combined with random samples, wherein positive samples include pixels belonging to the same class but predicted incorrectly and random pixels of the same class, and negative samples comprise pixels belonging to other classes but predicted incorrectly and random pixels of other classes. This sampling strategy fails to focus attention on the boundary information of the polyps and focuses on the segmentation of the overall target, which leads to insufficient sensitivity of the model to the polyp boundary information, thus affecting the segmentation effect. Hence, further research and promotion of polyp segmentation methods based on CIP-SCL are warranted.

To address the aforementioned challenges, this paper introduces a Strong-Guided Pixel Supervised Contrast (SGP-SCL) approach for polyp segmentation. The key features of this method are outlined as follows: (1) A novel strong-guided pixel-wise supervised contrastive learning method is presented. Leveraging ground truth information for each pixel, this method explores semantic relationships among the polyp's uncertain region, foreground region, and background region. (2) A strong-guided sampling strategy is proposed to tackle the issue of fuzzy polyp boundaries. By selecting reliable positive and negative samples, along with sampling anchor points from uncertain regions, this strategy establishes strong-guided learning objectives for the following contrastive learning. (3) A dynamic symmetric contrastive learning loss function is proposed, which symmetricly conducts contrastive learning for anchor points in the foreground region, and anchor points in the background region. Meanwhile, it utilizes a dynamic weight factor to balance supervision between the main loss function and the auxiliary loss function. (4) We validate our approach by performing experiments on widely recognized polyp segmentation datasets, including CVC-300, CVC-ClinicDB, Kvasir, CVC-ColonDB, and ETIS-LaribPolypDB, and achieve remarkable results.

II. Methods

A. Overall Architecture

Our method primarily emphasizes the sampling strategy and the construction of SGP-SCL. Our contrastive loss is inspired by CIP-SCL, and the specific formula for the contrast loss (cf.Eq(1)) in CIP-SCL is as follows:

$$L_i^{NCE} = \frac{1}{|P_i|}\sum_{i^+ \in P_i} -\log \frac{\exp(i \cdot i^+ / \tau)}{\exp(i \cdot i^+ / \tau) + \sum_{i^- \in N_i} \exp(i \cdot i^- / \tau)} \quad (1)$$

where, L^{NCE} called InfoNCE , is a popular loss function for contrastive learning [12], P_i and N_i denote the pixel embedding colections of positive and negative samples of pixel i, respectively, and τ >0 is a temperature hyper-parameter.

Specifically, as depicted in Figure.1, the proposed method is based on the encoder-decoder segmentation network. For an input image I, upon entering the segmentation network, the saliency map (m) is generated. Subsequently, m enters two branches for supervised training. In the first branch, m undergoes segmentation head to produce the prediction map, followed by the application of the cross-entropy loss. The specific operation of the segmentation head is mapping the feature map into a prediction map. In the second branch, employing the strong-guided sampling strategy, m selects positive and negative samples along with anchors, which are then projected into the embedding space. Ultimately, a dynamic symmetric contrastive loss is executed.

B. Strong-Guided Sampling Strategy

When applying the contrast learning method to polyp segmentation images, three challenges emerge: (1) Reduced sensitivity to small targets. When the segmentation target is small, the model's sensitivity to the foreground diminishes. Additionally, the target size may further decrease or disappear

during the downsampling across multiple network layers, leading to a failure to obtain samples from foreground class. (2) Imbalance in sampling positive and negative samples. Due to the small proportion of foreground areas and their similar features, only a limited number of positive samples are required for the network to learn foreground features. In contrast, the background region, with its larger proportion and diverse features, demands a substantial number of negative samples for the network to effectively distinguish between positive and negative samples. (3) Challenges with the uncertain region.The uncertain region in polyp images (which does not belong to the positive and negative samples) is the key reason for the low segmentation performance, and the previous methods applied to semantic segmentation failed to solve this problem.

In response to these challenges, this paper introduces a strong-guided strategy. Specifically, leveraging the encoder-decoder segmentation backbone network denoted as $f(*)$, the feature map F_l (*cf.*Eq(2)), which includes the last layer feature map, is extracted before each upsampling operation in the decoder of $f(*)$ is extracted for feature sample selection.

$$F_l = f(I), l \in [1,2,3,4] \tag{2}$$

Where l represents the label output of the four levels of the decoder, and I represents the input image. Moreover, the obtained F_l are all passed through the classifier to obtain the saliency map m_l (*cf.*Eq(3)).

$$m_l = Classifer(F_l) \tag{3}$$

Furthermore, in order to obtain the foreground region $m_{l,f}$ and the background region $m_{l,b}$ respectively in the saliency map m_l, the ground truth label $\hat{y}$ is used to divide the saliency map m_l into $m_{l,f}$ and $m_{l,b}$ (*cf.*Eq(4)).

$$m_{l,f} = m_l \cdot \hat{y},\ m_{l,b} = m_l \cdot (1-\hat{y}) \tag{4}$$

Sampling positive and negative samples. Based on the obtained $m_{l,f}$ and $m_{l,b}$, an active sampling strategy is constructed. Firstly designs a lower bound for positive samples collection, which aims to make the selected samples collection include as many foreground sample points with high confidence as possible. Similarly, designs the upper bound of negative samples collection, which aims to make the selected sample collection include as many background sample points with high confidence as possible. By selecting $m_{l,f}$ and $m_{l,b}$, the most reliable samples can be filtered, preventing unreliable samples from affecting the judgment of foreground and background regions. Specifically, suppose that the collection of selected positive samples is S_P (*cf.*Eq(5)), and the collection of selected negative samples is S_N *cf.*Eq(6).

$$S_P = \begin{cases} 1, m_{l,f} \geq \alpha \\ 0, other \end{cases} \tag{5}$$

$$S_N = \begin{cases} 1, m_{l,b} \leq 1-\alpha \\ 0, other \end{cases} \tag{6}$$

Where, a represents the selected probability value, typically falling within the range of $a \in [0.5,1]$. In the background region, given that the highest confidence often corresponds to the lowest confidence in the saliency map, 1-a is employed during selection. In this study, a is set to 0.95, ensuring both the quality of the selected samples, and the balanced of positive and negative samples. Following the selection of positive and negative samples, two reliable sample collections, S_P and S_N, are obtained. However, the prediction of uncertain regions remains indistinct, often resulting in the challenge of blurred edges in segmentation.

Sampling anchors. This paper selects the anchors at the boundary between positive and negative samples, and allows positive and negative samples to guide these critical anchors during model updates. The paper utilizes the intermediate probability value (0.5) from the saliency map as the critical point, where the uncertain region samples on both sides of the critical point are most abundant. Specifically, it is assumed that the anchors in the uncertain region are divided into two collections, A_P and A_N (*cf.*Eq(7)), where A_P represents the the collecion closer to the positive samples, and A_N represents the collecion closer to the negative samples.

$$\begin{cases} A_P, m_{l,f} \in [0.5, \theta] \\ A_N, m_{l,b} \in [1-\theta, 0.5] \\ 0, other \end{cases} \tag{7}$$

where θ represents a boundary value, we choose θ = 0.6, in order to select a moderate number of anchors from uncertain region.

Following the active sampling strategy, the samples are categorized into four sets, namely S_P, S_N, A_P and A_N. A_P and A_N represent the anchors of the uncertain region, which blur the boundary of the segmentation object, resulting in the degradation of segmentation performance. Leveraging the two reliable collections of positive and negative samples S_P and S_N, the model is proactively guided to effectively partition the samples within the uncertain region, thereby enhancing the segmentation performance of the boundary. Specifically, this paper proposes a dynamic symmetric contrastive loss to achieve S_P and S_N guided models to alleviate the issue of fuzzy boundaries.

C. Dynamic Symmetric Contrast Loss

Positive and negative samples (certain regions) are used to guide anchors (uncertain regions) for contrastive learning, so that the network can learn how to better distinguish the same class features and different class features. For the foreground, anchor $i \in A_P$, positive sample $i^+ \in S_P$, and negative sample $i^- \in S_N$. For the background, anchor $i \in A_N$, positive samples $i^+ \in S_N$, and negative samples $i^- \in S_P$. To this end, we construct a symmetric contrastive loss (*cf.*Eq.(8)).

$$L^{sym} = \begin{cases} \dfrac{1}{S_P} \sum\limits_{i^+ \in S_P} -\log \dfrac{\exp(i \cdot i^+)}{\exp(i \cdot i^+) + \sum_{i^- \in S_N} \exp(i \cdot i^-)}, i \in A_P \\ \dfrac{1}{S_N} \sum\limits_{i^+ \in S_N} -\log \dfrac{\exp(i \cdot i^+)}{\exp(i \cdot i^+) + \sum_{i^- \in S_P} \exp(i \cdot i^-)}, i \in A_N \end{cases} \tag{8}$$

Moreover, to ensure a balanced representation ability across different stages of model updates, we introduce a dynamic weighting factor λ. The value of λ is kept small in the early stages of model updates, and begins to increase during the middle and late stages. This choice is informed by the observation that, in the initial stages of model updates, the model's representation ability for the data is not well-established, and an excessively strong L^{sym} may impede the learning process. The overall loss function as follows:

$$L = L^{ce} + \lambda L^{sym} \tag{9}$$

where L^{ce} is the cross-entropy loss function and λ is the dynamic adjustment factor.

L^{sym} is dynamically introduced to co-supervise training after a period of L^{ce}. This is aimed at directing the network's primary focus towards accurately classifying pixels during the initial stage. As training epochs progress, the network's feature expression ability improves. Subsequently, L^{sym} is incorporated, with its weight gradually increased, to resegment the uncertain region. This process enhances the network's refined segmentation ability for polyp boundaries.

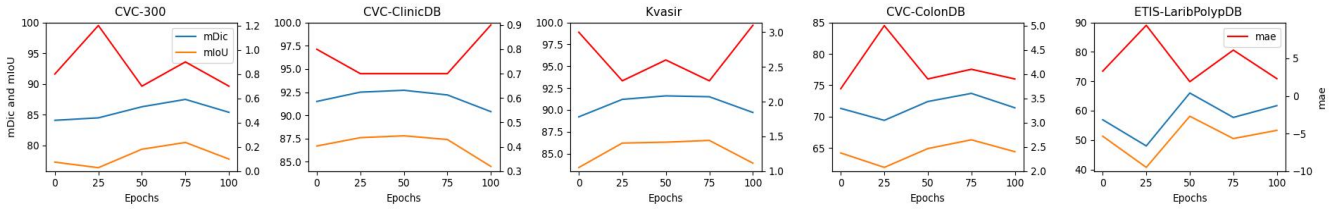

Figure.2. Comparison of adding SGP-SCL at different starting epochs.

TABLE I. COMPARISON OF DIFFERENT CONTRAST LEARNING METHODS

Method	CVC-300			CVC-ClinicDB			Kvasir			ETIS-LaribPolypDB		
	mDic	*mIoU*	*mae*	*mDic*	*mIoU*	*mae*	*mDic*	*mIoU*	*mae*	*mDic*	*mIoU*	*mae*
UCACNet	84.3	77.5	0.9	92.6	88.1	0.8	90.7	85.3	2.4	55.1	48.7	7.5
UCACNet + CIP-SCL	84.7	77.3	0.8	89.8	83.4	1.0	90.3	84.7	2.5	50.8	45.6	1.7
UCACNet + SGP-SCL	**86.3**	**79.4**	**0.7**	**92.7**	87.8	**0.7**	**91.6**	**86.3**	2.6	**66.0**	**58.1**	**1.9**

TABLE II. PERFORMANCE OF DIFFERENT Λ IMPLEMENTATIONS

λ	CVC-300			CVC-ClinicDB			Kvasir			CVC-ColonDB			ETIS-LaribPolypDB		
	mDic	*mIoU*	*mae*	*mDic*	*mIoU*	*mae*	*mDic*	*mIoU*	*mae*	*mDic*	*mIoU*	*mae*	*mDic*	*mIoU*	*mae*
0	84.3	77.5	0.9	92.6	88.1	0.8	90.7	85.3	2.4	74.5	67.3	4.1	55.1	48.7	7.5
5×10^{-3}	83.8	75.9	1.5	91.0	85.3	0.9	90.2	84.9	2.8	69.8	61.6	4.7	50.9	42.7	9.7
10^{-4}	**84.8**	76.7	1.0	89.4	84.6	0.8	89.1	83.3	3.0	71.3	63.7	4.3	51.5	43.9	**4.9**
5×10^{-4}	**86.1**	**77.9**	**0.8**	91.5	86.8	0.8	89.5	84.4	2.9	72.6	64.7	4.1	**59.2**	**51.0**	**2.4**
10^{-5}	**86.7**	**79.2**	0.9	92.4	88.1	0.8	**91.3**	**86.3**	**2.3**	74.1	66.9	**4.0**	**55.6**	**48.9**	**1.0**
5×10^{-5}	**86.3**	**79.4**	**0.7**	**92.7**	87.8	**0.7**	**91.6**	**86.3**	2.6	72.4	64.9	**3.9**	**66.0**	**58.1**	**1.9**
10^{-6}	**86.7**	**79.2**	0.9	92.4	88.1	0.8	**91.3**	**86.3**	**2.3**	74.1	66.9	**4.0**	**55.6**	**48.9**	**1.0**
5×10^{-6}	**87.7**	**81.1**	**0.8**	92.3	87.3	**0.7**	**91.2**	**86.2**	**2.1**	71.4	64.3	5.4	53.5	46.2	**3.9**
10^{-7}	**88.4**	**81.1**	0.9	91.0	86.4	0.8	**91.3**	**86.5**	2.4	74.3	66.4	4.2	**66.0**	**57.4**	**3.2**

TABLE III. QUANTITATIVE SEGMENTATION RESULTS

Method	CVC-300			CVC-ClinicDB			Kvasir			CVC-ColonDB			ETIS-LaribPolypDB		
	mDic	*mIoU*	*mae*	*mDic*	*mIoU*	*mae*	*mDic*	*mIoU*	*mae*	*mDic*	*mIoU*	*mae*	*mDic*	*mIoU*	*mae*
UNet	80.0	70.8	1.5	89.2	83.1	1.4	87.7	80.8	3.5	64.8	56.9	4.6	50.1	42.4	3.3
UNet++	81.0	72.4	1.4	88.3	83.7	1.3	88.2	81.7	3.3	65.2	56.7	4.9	43.7	36.5	4.8
PraNet	87.4	80.1	0.8	91.0	86.2	0.8	91.4	86.4	2.4	74.6	66.6	3.7	65.9	59.2	3.2
UCACNet	84.3	77.5	0.9	92.6	88.1	0.8	90.7	85.3	2.4	74.5	67.3	4.1	55.1	48.7	7.5
UCACNet+SGP-SCL	**86.3**	**79.4**	**0.7**	**92.7**	87.8	**0.7**	**91.6**	**86.3**	2.6	72.4	64.9	**3.9**	**66.0**	**58.1**	**1.9**
LDNet	83.5	76.3	1.2	90.5	85.0	1.1	91.2	85.7	2.4	76.2	68.7	3.6	66.0	58.0	2.9
LDNet+SGP-SCL	**85.4**	**78.6**	**1.1**	**91.9**	**87.1**	0.8	90.9	85.4	2.6	**76.4**	**69.1**	3.7	**68.3**	**60.8**	**2.1**

III. EXPERIMENT

A. Dataset

We use the same training data as [11] to make a fair comparison. That is, 550 images from CVC-ClinicDB [13], and 900 images from kvasir [14] were selected as training set. The remaining 62 images in CVC-ClinicDB, and 100 images in kvasir served as the test set. All images from the CVC-300, CVC-ColonDB [15], and ETIS-LaribPolypDB [16] datasets were used as test sets to test the network's generalization performance on previously unseen datasets.

CVC-300, is a dataset from EndoScene [17]. EndoScene contains 912 images of 44 colonoscopy sequences from 36 patients. The CVC-300 dataset is the 60 images in EndoScene and their corresponding real labels.

CVC-ClinicDB, also known as CVC-612. The dataset contains 612 images and corresponding real labels with image sizes of 384 × 288.

Kvasir, an endoscopic dataset for pixel-level segmentation of colon polyps, consists of 1000 images of gastrointestinal polyps and their corresponding segmentation masks, which are personally labeled and verified by senior gastroenterologists.

CVC-ColonDB, was derived from 15 different colonoscopy sequences, and 380 images were sampled from these sequences.

ETIS-LaribPolypDB, a dataset containing 196 images collected from 34 colonoscopy videos. The size of the image is, which is the largest in other data sets. The polyps in this dataset are all small and difficult to find, making this dataset even more challenging.

B. Experimental details

Our data enhancement is consistent with [11], using the Adam [18] optimizer, and sets the initial vector to 10^{-4}, vector polynomial attenuation factor [19] to $(1-(epoch/epoch_{\max})^{0.9})$, training epoch to 240, input image compression to 352 × 352, training batch size to 16. In particular, the test results are calculated by upsampling the prediction map of size 352 × 352 back to the original size.

C. Ablation experiment

In this study, UCACNet serves as our foundational segmentation network, and we conduct a comprehensive ablation experiment with 240 epochs for each experiment while maintaining constant values for other hyperparameters. The evaluation metrics employed in this analysis encompass mDice, mIoU, and MAE. These metrics offer insights into the accuracy, degree of overlap, and overall quality of the segmentation results generated by each method.

Comparison of different contrast learning methods. In this paper, we initially investigated the efficacy of our SGP-SCL, and CIP-SCL in polyp image segmentation. Ablation experiments were conducted on the CVC-300, CVC-ClinicDB, Kvasir, and ETIS datasets. As depicted in Table I, when considering SGP-SCL in segmentation network learning—utilizing reliable positive and negative samples to guide anchors learning—it demonstrates notable performance improvements across the four datasets compared to the "baseline (no contrast)" approach. Particularly noteworthy is the substantial enhancement on the most challenging ETIS-LaribPolypDB dataset, where the mIOU index increases by 10.9%. Conversely, the CIP-SCL method, which samples difficult and random samples without targeted guided learning, does not exhibit significant performance improvement, and even reduces performance on most datasets.

Comparison of different starting-epochs. In the method proposed in this paper, the choice of the number of starting epochs significantly influences the model's performance. As illustrated in Figure.2, it shows that adding auxiliary functions to the model at 50 epochs yields optimal test results in more datasets, and provides the best overall performance on all datasets, without the occurrence of notably poor performance on any specific dataset. Notably, adding auxiliary functions either too late or too early leads to suboptimal test results. This phenomenon can be attributed to the fact that the model's representation ability is not fully established with a small number of training epochs, and the introduction of L^{sym} may disrupt the process of representation learning. Conversely, with more training epochs, the representation power of the model approaches steady levels and is therefore not susceptible to L^{sym}.

Comparison of different dynamic weighting factors. To investigate the impact of the loss weight of L^{sym}, we varied the sizes of the weighting factors (λ), and observed the corresponding changes in model performance. As shown in Table II, when λ is excessively large, the model becomes overly influenced by L^{sym}, hindering the establishment of basic data representation ability. When $\lambda > 10^{-5}$, the model can show better results on most of the data sets. In this paper, considering the performance on all the data sets, $\lambda = 5\times10^{-5}$ is finally selected as the weighting factor of the proposed method.

D. Comparative experiments

The experiments on the polyp image segmentation task used five different datasets including CVC-300, CVC-ClinicDB, Kvasir, CVC-ColonDB and ETIS-LaribPolyPDB to evaluate the performance of different segmentation methods. Experimental results have been summarized in Table III. Which compares the segmentation results of the classical semantic segmentation methods UNet, UNet++, the leading polyp segmentation method PraNet, UCACNet, LDNet [56], and experiments combined with the proposed method in this paper UCACNet + SGP-SCL and LDNet + SGP-SCL.

It is obvious that the polyp segmentation method combined with SGP-SCL can obtain more accurate segmentation results in the face of most challenges. On the CVC-300 and ETIS-LaribPolypDB dataset, the segmentation effect of the SGP-SCL method is more significant, providing better test results for the performance of the model. In particular, UCACNet +

SGP-SCL, which was improved by 9.4% on the mIou index on the ETIS dataset over methods that did not use SGP-SCL.

In addition, the visualization results of some of the experiments are shown in Figure.3. For the CVC-300 dataset, the polyp target accounted for a small proportion of the area in the colonoscopy image, and LDNet showed the phenomenon of redundant recognition, but was solved when the SGP-SCL method was introduced. In the CVC-ClinicDB dataset, the polyp target accounts for a large proportion of the area in the colonoscopy image, most models can segment it well. For the Kvasir dataset, the intestinal environment is complex, which leads to a large number of incorrectly segmented regions for the target in the UACANet and LDNet models. After the introduction of the SGP-SCL method, the segmentation accuracy of the model has been significantly improved. In the CVC-ColonDB dataset, the model with the addition of SGP-SCL method achieves higher segmentation performance, which effectively supplements the original missing segmentation. In the ETIS-LaribPolypDB dataset, the model with the SGP-SCL method can also effectively reduce the missegmentation phenomenon and make up for the missing segmentation area. In summary, it is confirmed that the SGP-SCL method can indeed improve the segmentation accuracy of polyp images of the model.

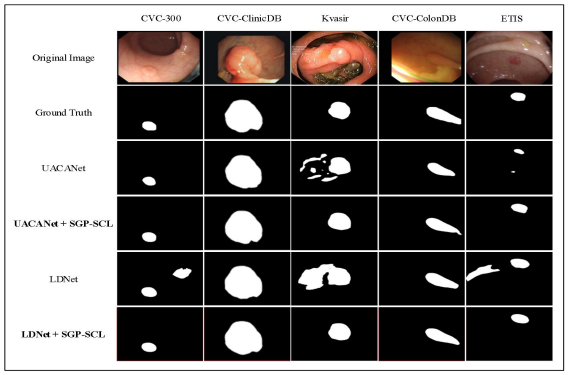

Figure.3. Performance visualization.

IV. Conclusion

In this paper, we propose a novel pixel-wise supervised contrastive learning method called SGP-SCL, which enhances polyp boundary by leveraging reliable positive and negative samples guide uncertain anchors learning. Additionally, the dynamic symmetric contrast loss enhances the learning effect of the segmentation network through contrastive learning of the foreground and background. In a series of contrast and ablation experiments, our method demonstrates promising results.However, our method also deserves further discussion and research, and we hope to achieve better results in a wider data set in the future.

Acknowledgment

This work was supported in part by the Natural Science Foundation of Guangdong Province under Grant 2022A1515010130 and Grant 2023A1515011793.

References

[1] Smith RA, Fedewa S, Siegel R. Early colorectal cancer detection-Current and evolving challenges in evidence, guidelines, policy, and practices. Adv Cancer Res, 2021, 151: 69-107.

[2] Gupta S. Screening for colorectal cancer[J]. Hematology/Oncology Clinics, 2022, 36(3): 393-414.

[3] Hamida A B, Devanne M, Weber J, et al. Deep learning for colon cancer histopathological images analysis[J]. Computers in Biology and Medicine, 2021, 136: 104730.

[4] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[5] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation[M]//Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham, 2018: 3-11.

[6] Fan D P, Ji G P, Zhou T, et al. Pranet: Parallel reverse attention network for polyp segmentation[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2020: 263-273.

[7] Fang Y, Chen C, Yuan Y, et al. Selective feature aggregation network with area-boundary constraints for polyp segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 302-310.

[8] Zhong J, Wang W, Wu H, et al. Polypseg: An efficient context-aware network for polyp segmentation from colonoscopy videos[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2020: 285-294.

[9] Guo Q, Fang X, Wang L, et al. Polyp Segmentation of Colonoscopy Images by Exploring the Uncertain Areas[J]. IEEE Access, 2022.

[10] Yue G, Han W, Jiang B, et al. Boundary Constraint Network with Cross Layer Feature Integration for Polyp Segmentation[J]. IEEE Journal of Biomedical and Health Informatics, 2022.

[11] Kim T, Lee H, Kim D. Uacanet: Uncertainty augmented context attention for polyp segmentation[C]//Proceedings of the 29th ACM International Conference on Multimedia. 2021: 2167-2175.

[12] Wang W, Zhou T, Yu F, et al. Exploring cross-image pixel contrast for semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 7303-7313.

[13] Bernal J, Sánchez J, Vilarino F. Towards automatic polyp detection with a polyp appearance model[J]. Pattern Recognition, 2012, 45(9): 3166-3182.

[14] Jha D, Smedsrud P H, Riegler M A, et al. Kvasir-seg: A segmented polyp dataset[C]//MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5 −8, 2020, Proceedings, Part II 26. Springer International Publishing, 2020: 451-462.

[15] Bernal J, Sánchez F J, Fernández-Esparrach G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians[J]. Computerized medical imaging and graphics, 2015, 43: 99-111.

[16] Silva J, Histace A, Romain O, et al. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer[J]. International journal of computer assisted radiology and surgery, 2014, 9: 283-293.

[17] Vázquez D, Bernal J, Sánchez F J, et al. A benchmark for endoluminal scene segmentation of colonoscopy images[J]. Journal of healthcare engineering, 2017 : 9.

[18] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).

[19] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017).

[20] Zhang R, Lai P, Wan X, et al. Lesion-aware dynamic kernel for polyp segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022: 99-109