HSNet+: Enhancing Polyp Segmentation with Region-wise Loss

Pietrobon Andrea and Biffis Nicola

Department of Engineering Information, University of Padua

Abstract

This research will show an innovative method useful in the segmentation of polyps during the screening phases of colonoscopies with the aim of concretely helping doctors in this task. To do this we have adopted a new approach which consists in merging the hybrid semantic network (HSNet) architecture model with the Reagion-wise(RW) as a loss function for the backpropagation process. In this way the bottleneck problems that arise in the systems currently used in this area are solved, since thanks to the HSNet it is possible to exploit the advantages of both the Transformers and the convolutional neural networks and thanks to the RW loss function its capacity is exploited to work efficiently with biomedical images. Since both the architecture and the loss function chosen by us have shown that they can achieve performances comparable to those of the state of the art working individually, in this research a dataset divided into 5 subsets will be used to demonstrate their effectiveness by using them together .

Introduction

In the field of biomedicine, quantitative analysis requires a crucial step: image segmentation. Manual segmentation is a time-consuming and subjective process, as demonstrated by the considerable discrepancy between segmentations performed by different annotators. Consequently, there is a strong interest in developing reliable tools for automatic segmentation of medical images.

The use of neural networks to automate polyp segmentation can provide physicians with an effective tool for identifying such formations or areas of interest during clinical practice. However, there are two important challenges that limit the effectiveness of this segmentation:

1. Polyps can vary significantly in size, orientation, and illumination, making accurate segmentation difficult to achieve.
2. Current approaches often overlook significant details such as textures.

To obtain precise segmentations in medical image segmentation, it is crucial to consider class imbalance and the importance of individual pixels. By pixel importance, we refer to the phenomenon where the severity of classification errors depends on the position of such errors.

Current approaches for polyp segmentation primarily rely on convolutional neural networks (CNN) or Transformers, so to overcome the mentioned challenges, this research proposes the use of a hybrid semantic network (HSNet) that combines the advantages of Transformer networks and convolutional neural networks (CNN), along with regional loss (RW). Thanks to this loss function we can simultaneously takes into account class imbalance and pixel importance, without requiring additional hyperparameters or functions, in order to improve polyp segmentation.

In this study, we examine how the implementation of regional loss (RW) affects polyp segmentation by applying it to a hybrid semantic network (HSNet).

Methods

HSNet - Hybrid Semantic Network

The HSNet is a hybrid semantic network (HSNet) that combines Convolutional Neural Network (CNN) and Transformer; in this way it can exploit both the CNNs success in computer vision tasks and the Transformers excellence in capturing long-range dependencies and global contextual features.

In particular, the proposed HSNet incorporates a cross-semantic attention module (CSA) to filter out noise and bridge the semantic gap between the encoder and decoder stages, it also includes a dual-branch hybrid semantic complementary module (HSC) that combines interactive features from Transformers and local details perceived by CNNs to restore appearance details of polyps. Furthermore, a multi-scale prediction module (MSP) with learnable weights can integrates stage-level predictions effectively.

RW - Region-Wise Loss

The region-wise loss is a loss function used in semantic segmentation tasks defined as the sum of the element-wise multiplication between the softmax probability values and a region-wise map $Z = (z)_{ik} = [z_1, ..., z_N]$ computed based on the ground truth and independent of the network's parameters:

$$L_{RW} = \sum_{i=1}^{N} y_i^T z_i = \sum_{i=1}^{N} \sigma(\psi_i)^T z_i$$

where ψ_i denote the unormalized prediction of a ConvNet, y_i its softmax normalized values, z_i is the map value and σ is the softmax function.
Region-wise loss yields the following gradients with respect to the unnormalized prediction of a ConvNet:

$$\frac{\partial L_{RW}}{\partial \psi_{ik}} = \frac{\partial L_{RW}}{\partial y_i} \frac{\partial y_i}{\partial \psi_{ik}} = y_{ik} \sum_{l=1;l \neq k}^{K} y_{il}(z_{ik} - z_{il})$$

The fundamental aspect of this loss function is its versatility in reformulating other loss functions and its ability to penalize pixels based on their class and location, providing a more flexible approach compared to traditional loss functions like Cross entropy.
An other principle, for rectifying RW maps to ensure optimization stability, is the Rectified Region-wise map (RRW), which is a normalized version of RW-Boundary maps used to handle class imbalance and to enable transfer learning, computing the map during the optimization and decreasing/increasing the contribution of certain images to the loss.
Thanks to these aspects this function works so well to operate with biomedical-type images.

Implementation

The implementation of the hybrid semantic network (HSNet) with the RW loss function for polyp segmentation involved several key steps. Firstly, we started from the HSNet architecture which integrates CNN and Transformer components in this way: the encoder-decoder structure was designed incorporating convolutional layers, attention mechanisms, and skip connections; while the cross-semantic attention module (CSA) and the dual-branch hybrid semantic complementary module (HSC) are implemented as intermediate transition modules to address semantic gaps and restore appearance details.

To facilitate model training, we used polyp segmentation datasets, including Kvasir-SEG, ClinicDB, ColonDB, Endoscene, and ETIS; moreover data preprocessing was performed, including image resizing, normalization, and augmentation techniques.

During training, we employed the RW loss function and the model was optimized using the AdamW optimization algorithm, with carefully selected hyperparameters such as learning rate, batch size, and number of epochs. We also incorporated evaluation metrics like Dice coefficient to assess the model's performance.

Within the training loop, we iteratively fed the training dataset through the HSNet model, calculating the RW loss and updating model parameters via backpropagation. We precisely fine-tuned the model's hyperparameters, including CSA and HSC configurations, to achieve optimal segmentation results and to ensure reproducibility, we carefully documented the specific CNN and Transformer architectures, layer configurations, and hyperparameters used in the implementation.

Following training we evaluated the HSNet model quantifying segmentation accuracy using metrics like Dice coefficient and we visually compared the predicted segmentation masks with ground truth annotations for qualitative analysis.

To assess the generalization capabilities of the trained HSNet model, we applied it to unseen data or separate test datasets and we calculated evaluation metrics on the test set, comparing them with the results obtained on the training set to measure the model's performance.

Overall, this implementation of HSNet with the RW loss function provides an effective framework for polyp segmentation, addressing the challenges of semantic gaps, detail recovery, and global context modeling.

Results

As previously mentioned, the model was trained using a dataset made up of 6 categories of images containing different types of polyp, divided into training and evaluations, while a similar dataset with new images was provided for the testing phase. Various tests were carried out in order to identify the correct hyperparameters to be used to make the model working well, taking in to account the hardware resources at our disposal, thus identifying the following parameters:

- learning rate: 0.0001
- training size: 160
- batch size: 24
- number of epochs: 50

The best results obtained are those that we can see both in Table 1 and in Figure 1, showing a model that is not yet perfect and cannot reach high Dice levels but which with other training processes could achieve the expected results.

Table 1: *Table showing the Dice values for each set of images every 10 epochs*

epoch	10	20	30	40	50
CVC-300	0.49	0.38	0.69	0.55	0.33
CVC-ClinicDB	0.60	0.58	0.71	0.58	0.39
Kvasir	0.68	0.69	0.73	0.63	0.53
CVC-ColonDB	0.35	0.31	0.51	0.41	0.31
ETIS	0.41	0.37	0.51	0.37	0.30
Test	0.43	0.40	0.57	0.45	0.34

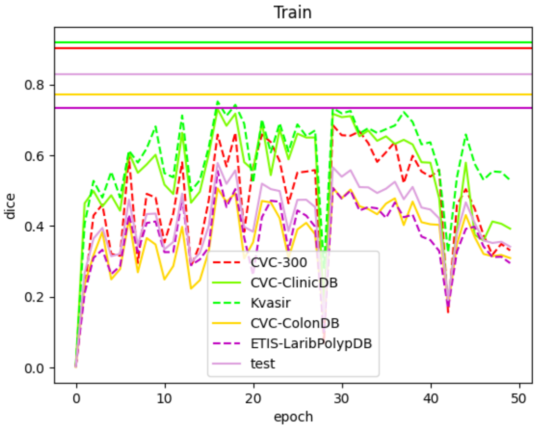

Figure 1: *The graph shows the trend of the value of Dice on all 6 types of images based on the current epoch. As we can see, the model obtains higher values with certain sets of images and lower ones with others.*

In any case, as we can see from Figure 2 and Figure 3, using it we can obtain quite precise masks.

Conclusion

In conclusion, this paper presents a comprehensive investigation into the problem of polyp segmentation in endoscopic imaging. Through the implementation of the hybrid semantic network (HSNet) combined with the RW loss function, we have addressed key challenges in the field, including semantic gaps, detail recovery, and global context modeling.

Our experimental results on multiple challenging polyp segmentation datasets, including Kvasir-SEG, ClinicDB, ColonDB, Endoscene, and ETIS, demonstrate the effectiveness of HSNet compared to existing state-of-the-art models. However we are aware that the results obtained have significant margins of improvement, due to the fact that we had limited hardware resources at our disposal, which prevented us from fully leveraging the model.

The incorporation of both CNN and Transformer components in HSNet allows for the separate capture of local and long-term features, leading to improved segmentation accuracy, while the utilization of the RW loss function further contributes to the optimization stability of the model and enhances its convergence during training.

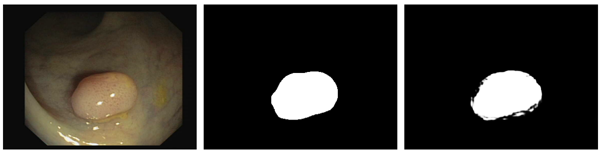

Figure 2: *From left to right, in order: input image, true mask, predicted mask.*

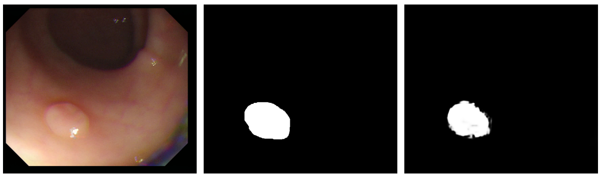

Figure 3: *From left to right, in order: input image, true mask, predicted mask.*

Moreover, the introduction of the cross-semantic attention module (CSA) and the dual-branch hybrid semantic complementary module (HSC) in HSNet addresses the challenges of semantic gaps and detail restoration, resulting in more precise segmentation masks with restored appearance details.

Our study not only contributes to the advancement of polyp segmentation techniques but also sheds light on the potential benefits of hybrid models combining CNN and Transformer architectures in medical image analysis tasks. Finally, the proposed HSNet framework can serve as a foundation for future research and can be extended to other medical imaging applications beyond polyp segmentation.