Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

GUNet: Duplex Global Graph Relationship Mining for Polyp Segmentation

Wumiao Xie, Kaicheng Frank Yang

BASIS International School Guangzhou, Guangzhou, 510663, China

Abstract. Colorectal Cancer is one of the most common cancers worldwide, and as the probability of cure decreases as the stages progress, the early-stage detection of possible indications of colorectal cancer (i.e., polyps) during colonoscopy exams can contribute greatly to its prevention and curation. Since some of the indications of colorectal cancer are hard to identify by the human eye, it becomes crucial to develop an algorithm that can assist in the early identification of abnormal tissue walls that may be signs of early-stage cancer. This paper proposes the GUNet, where duplex global graph relationships are mined through graph convolution operations. The core of this framework is the extraction and utilization of both spatial and semantic information from diverse encoded feature maps to acquire contextual semantic information, done so to construct a fuller and more diverse global representation. We also train the GUNet on a combination of three public datasets to enhance its generalization abilities. Experiments show that the GUNet can serve as a competitive alternative to the U-Net, achieving a 64.47% IOU on the combined dataset.

Keywords: Convolutional Neural Network; Semantic Segmentation; Graph Convolutional Network.

1. Introduction

Colorectal Cancer (CRC) is the third most common cancer worldwide and the second-most prevalent cause of all cancer-related deaths. The overall five- year survival rate for colon cancer is around 63%, and that of rectal cancer around 68% [1]. As the survival rate of CRC declines substantially with the disease's aggravation, with early-stage survival rates reaching 90% while 5-year distant-stage dropping to as low as 14%, early-stage diagnosis of precancerous anomalies becomes crucial in combating CRC. While colonoscopy has become the standard practice in clinics for the early detection and removal of polyps, due to the procedure's labor-intensiveness, the varying appearance of polyps, and the varying skill of the endoscopist, up to 25% of polyps may be missed during the process [2], [3]. Computer Aided Diagnosis (CAD) systems can contribute significantly to addressing this situation by assessing the colonoscopy results more thoroughly, eliminating observer variation, and lightening the bur-den of physicians [4]. Therefore, designing reliable and automatic segmentation methods is a necessity in achieving a more accurate early-stage diagnosis of CRC. (As shown in figure 1).

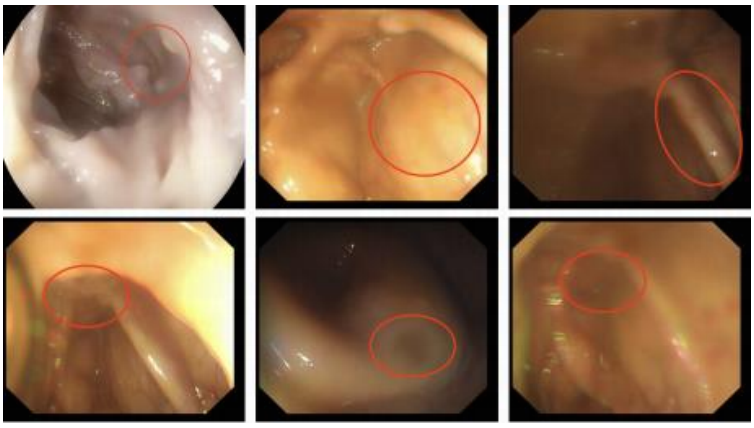

Figure 1. Some examples that indicate the inherent challenges when attempting to perform polyp segmentation, including but not limited to varying polyp sizes, blurred boundaries and images, and presence of extraneous materials.

In the past decades, automatic polyp segmentation has received increasing attention. Early works consisted of mostly hand-crafted methods, exploring and analyzing polyp characteristics from color distribution [5], textural properties [6], contour regions [7], protrusion measurements [8], etc. While these approaches can be effective within certain scenarios, they lead to a generally lackluster performance due to (i) inherent polyp qualities, their variations in appearance and their boundaries with mucosa; (ii) the dependency of hand- crafted methods on the ability and expertise of their designer(s); (iii) their focus on certain patterns causes them to ignore otherwise useful information, all of which limits their general performance.

More recently, approaches based on convolutional neural networks (CNNs) have gained much traction, showing obvious advantages over traditional methods with their ability to perform end-to-end, automatic feature extraction and integration to an unprecedented precision [8], [9]. Some important milestones have been made with general-purpose segmentation models: in Ref. [10], the fully convolutional network (FCN) is proposed; and in Ref. [11], an encoder-decoder structure for medical image segmentation is proposed as the U-Net.

Many follow-up works have built on the encoder-decoder architecture, updating the encoder and decoder module(s) [12], [13], introducing innovations based on skip connections [14], [15], [16] and attention mechanisms [17], [18] etc. While CNN-based methods undoubtedly show improved performance upon hand-crafted methods and do decrease expertise-dependency, the full utilization of global in-formation within polyp segmentation for the traditional encoder-decoder architecture still remains a challenge. While some previous works have attempted to address this issue with the use of recurrent neural networks (RNNs), their performance ultimately fell short due to inherent architectural difficulties and a lack of development in the spatial information handling crucial for polyp segmentation [19], [20].

In this paper, we propose a novel encoder-decoder architecture to address the aforementioned issue. To effectively construct a complete initial representation, we first combine the two bottom-most layers of the U-Net. Then, we establish two adjacency matrices to represent distinctive spatial and semantic information respectively, after which we implement two parallel graph-based convolutional modules to extract and construct two distinctive aspects of global information. We finally combine this to construct a full global representation complete with both spatial and semantic information. In addition, we combine three individual datasets into one to train on a larger quantity and variety of data. We show that our model effectively extracts global contextual semantic information through experimentation.

Our contributions can be summarized as follows:

1. We propose the GUNet, a novel encoder-decoder architecture that utilizes graph-based modules in the form of skip connections, extracting enhanced global information and decreasing the semantic gap between the encoder and the decoder.

2. We propose the use of two distinctive adjacency maps modeling both spatial and semantic relationships from within the feature map. This is utilized with two parallel graph convolution modules to extract the global semantic information from both, combining them into a fuller representation of both spatial and semantic relationships.

3. To make up for the lack of polyp segmentation training data, we have combined the three individual datasets of ColonCB, Kvasir-Seg, and ETIS into one dataset, providing a larger quantity of data for our network to train on, as well as a greater variety of data for better generalization.

The paper is organized as follows: In section 2, we present related works on the segmentation of colonoscopic polyps. In section 3, we describe our proposed approach with the GUNet. In section 4, we describe the dataset and metrics utilized and present an evaluation of experiments done on our model. In section 5, we conclude the paper, and in section 6, we acknowledge our contributors.

2. Related Works

Over the last few decades, automated polyp segmentation has become an active and notable topic for clinical and academic research, with considerable work being put into it in order to address the significant issue of CRC, achieving more efficient and effective results. In this section, we analyze some significant past works done in automated polyp segmentation from the following three aspects.

2.1. Methods Based on Hand-crafted Features

In the early development of automatic polyp segmentation, many works focused on the utilization of handcrafted features in their approach, the majority exploiting low-level image processing operations to effectively explore and analyze the characteristics of polyps. Karkanis et al. [5] utilized discrete wavelet decomposition being used across different color spaces. Hwang et al. [21] focused on the common shape of small polyps and proposed a technique combining ellipse fitting and curvature analysis. Vân Wijk et al. [22] measures the amount of protrusion within polyps in a scale-adaptive manner. Silva et al. [23] utilized a detection algorithm based on the circular shape of polyps, followed by the cooccurrence matrix. Mamanov et al. [24] classified frames based on their best-fit ball radius under the assumption that polyps are characterized by mostly round protrusions. Iwahori et al. [25] extracted polyp candidate regions using the Hessian filter, from which Histograms of Oriented Gradients (HOG) features are extracted. Bernal et al. [26] used valley depth energy maps (WM-DOVA maps) to integrate valley information, and S´anchez-Gonz´alez et al. [7] analyzed the appearance of contour regions through their shape, color, and curvature in order to be used for feature extraction.

While these handcrafted feature-based approaches did attract more attention to automatic polyp segmentation and can be effective within certain scenarios, these methods lead to a generally lackluster performance, with development being restricted by many differing factors. Among these include (i) inherent polyp qualities, i.e. their variations in appearance, such as in size, texture, and shape, as well as the ambiguous boundaries between them and mucosa; (ii) the dependency of hand-crafted methods on the ability and expertise of their designer; (iii) their focus on certain patterns otherwise useful information that do not cater to the situation to be ignored, limiting general performance.

2.2. General-Purpose Methods Based on Deep Learning

More recently, segmentation methods based on CNNs have gained much traction, receiving attention for their increased effectiveness in handling general and challenging scenarios compared to approaches based on handcrafted features. CNN-based methods can be roughly divided into two categories: methods that are general-purpose and methods that are specific-purpose.

General-purpose methods are, as their name suggests, general in their purpose, aiming to develop a universal network suitable for a wide range of different tasks. For example, long et al. [10] proposed the famous fully convolutional network (FCN), which replaced the traditional fully connected layers of CNNs with more convolutional layers, achieving state-of-the-art performance in multiple segmentation datasets. Evolving from the FCN, Ranneberger et al. [11], addressing the difficulty of medical image segmentation tasks, proposed the U- Net, which possesses a U-shaped encoder-decoder architecture heavily sampled by later works. Oktay et al. [17] directly builds off the encoder-decoder architecture and introduces attention gate modules, and Diakogianniset al. [27] utilized the architecture in combination with residual connections, Zhou et al. [14] introduced redesigned embedded skip connections to the original architecture and utilizes multi-depth U-Net's. Some works have already begun improving upon the global and local information integration. For example, Song et al. [28] introduced the global and local feature reconstruction modules to better integrate global context and up-sample local features.

2.3. Specific-Purpose Methods Based on Deep Learning

The specific-purpose methods aim to design a specific model targeted solely towards polyp segmentation, considering its unique characteristics that general- purpose methods would otherwise choose to ignore due to their infrequent appearance in other tasks yet lacking in the generalizability that general-purpose methods possess. Many works build off of general-purpose methods, making modifications to fit the specific task better. For example, Jha et al. [12] built off previous, general-purpose work by introducing multi-resolution dense skip connections and attention modules to the Res UNet architecture. Banik et al. [29] implements a fusion-based architecture that is based on the original CNN and utilizes a nascent pooling mechanism with a novel level-set method to suppress false positive regions. Sun et al. [30] introduced an altered architecture based on the U-Net, implementing the dilated convolution to improve feature representation and several post-processing strategies to smooth the segmentation boundaries. Patelet al. [31] proposed an enhanced, attention-based variant of the U-Net that utilizes new modules to enhance semantic information and focus on only finegrained features from the encoder. Jha et al. [13] built off of the encoder-decoder architecture to achieve faster processing speed while not sacrificing significant performance. Zhao et al. [32] introduced multi-scale subtraction operations to the U-Net architecture, reducing redundant information from the encoder.

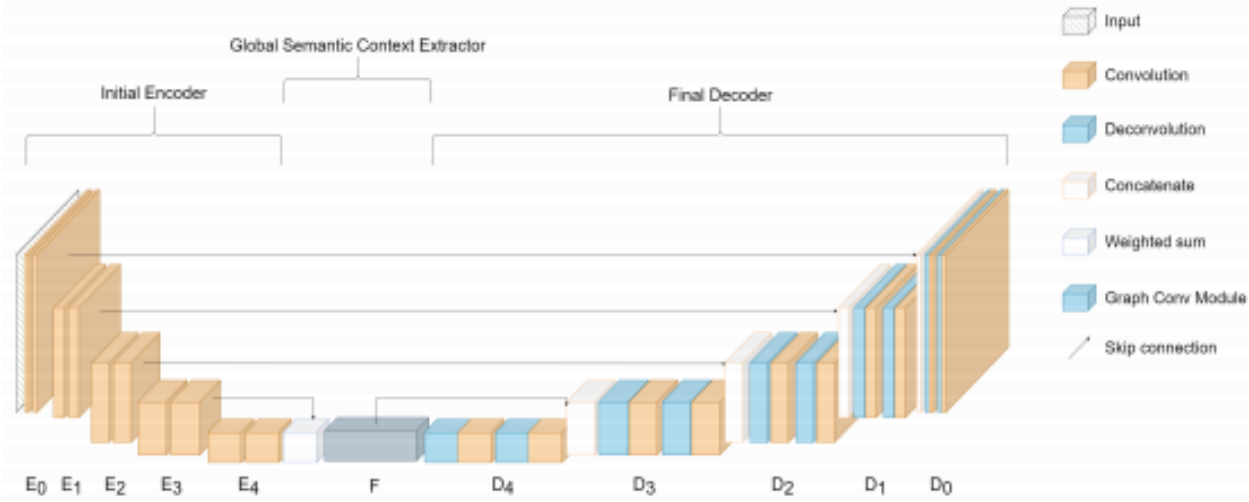

Figure 2. A schematic illustration of GUNet's architecture

Many recent works have placed a larger emphasis on expanding methods to exploit latent information, such as Mahmud et al. [16] integrating parallel dilated convolutions of different dilation rates to increase receptive area, and many focused especially on further exploiting the global (and local) information. Yinetal. [18] integrated cross-image contextual relations with the intent of enhancing feature selection, considering contextual correlation on top of internal correlation and achieving state-of-the-art. Srivasta et al. [33] employed multi-scale fusion, attention, and feature-selection operations designed to improve the architecture's generalizability across varying datasets with varying data distributions. Yue et al. [34] similarly focused on integrating multi-scale and multi-level features, yet more so leveraging boundary information to enhance the polyp representation. Yue et al. [35] employ an attention-based pyramid architecture with a simple context extraction module, which replaces the traditional convolutions using multiple receptive field sizes with parallel branches aimed to combine local and global information. Zhang et al. [36] introduce an adaptive context module, which adaptively selects local and global information regions based on the size of the polyp generated.

3. Method

3.1. Overview

Our model utilizes the traditional encoder-decoder architecture, and can be split up into three sections: the initial encoder from the traditional architecture; the global extractor, where the proposed global semantic context extractor module extracts and combines two distinctive levels of global relationships of a spatial semantic matrix and a semantic adjacency matrix (respectively); and lastly the final decoder, where the results from the initial and global extractors are inputted as skip connections to the traditional decoder. In fig.2, we provide a schematic illustration of the workflow of our architecture, completed in an end-to-end manner.

3.2. Initial Encoder

As shown in fig. 2, the traditional encoder-decoder architecture, specifically the original U-Net implementation, is adopted as our model's backbone. In the initial encoder section of the GUNet, we employfive encoder blocks, with each consisting of a 3×3 convolutional block, a batch normalization module, and a ReLU activation function repeated two times (except the first) and their configurations listed in the illustration. Specifically, given $\{X, Y\}_{n=1}^{N}$, $X \in R^{C\times H\times W}, Y \in R^{H\times W}$, where X denotes the input image, Y the ground truth, and N, n, C, W, H the data-set size, the current set index, the number of channels of X, the width of X, and the height of X respectively, we compute the side-outputs for the encoder, which we denote as E_i, with size: $64 \times 2^i \times \frac{H}{2^i} \times \frac{W}{2^i}$ $\forall i \in \{1, 2, 3, 4, 5\}$. We compute Ei as the following:

$$\begin{cases} E_i = \text{down}(E_{i-1}), i \in \{1, 2, 3, 4\} \\ E_i = \text{doubleConv}(E_{i-1}), i = 0 \end{cases} \tag{1}$$

Where doubleConv (·) denotes two convolution operations and down (·) de-notes the encoder blocks.

3.3. Global Extractor

The extraction of effective semantic information amongst features are cru-cial in achieving precise segmentation, and after the initial encoding, deeper layers with higher-level features are best suited for this task, possessing greater capabilities for the representation of semantic information. To perform this task, as part of the global extractor, we first combine the feature maps from the third and fourth layer of the encoder. Compared to the fourth layer, the third layer contains relatively more features with greater resolutions, and thus their features are complimentarily combined when the feature maps are merged.

$$F = w_a * E_3 + w_b * E4 \tag{2}$$

Here, weighted addition is used to combine the two feature maps, with ⇤–denoting element-wise multiplication and w_a and w_b respectively denoting the trainable weights for E_3 andE_4.

Through our research, we have discovered two distinctive strands of global information: spatial correlation, where relationships are encoded based on spatial proximity, and semantic correlation, where relationships are encoded based on semantic similarity. For the former, spatial relationships are useful as they i) help contextualize both local and global information of the pixels, and ii) discover spatial correlations and patterns within the data, enabling the model to perform precise space-related tasks such as locating polyp boundaries. When used together with the latter, which offers semantic global contextual information, the two compliments each other into building a fuller and more complete representation of global information. We incorporate both through utilizing two adjacency matrices, one based on spatial relationships, and the other semantic relationships. Specifically:

$$A_{jk}^{sp} = \begin{cases} 1, k \in Ne(j) \\ 0, otherwise \end{cases} \quad A_{jk}^{se} = d(f_j, f_k) \tag{3}$$

Where $j, k \in \left\{0, 1, 2, \cdots \frac{H}{16} \times \frac{W}{16} - 1\right\}$. Here, $Ne(j)$ denotes the neighbors of j in F, $f_{j\ \text{denotes}}$ the location j of the combined feature map which we have flattened pre-processing(f), and d (f_j, f_k) denotes the semantic similarity between f_j and f_k. Ford (f_j, f_k), we compute the semantic similarity using cosine similarity, which we have evaluated to be the most suitable option and can be expressed as the following:

$$d(f_j, f_k) = \frac{f_j \cdot f_k}{\|f_j\|_2 \|f_k\|_2} \tag{4}$$

Subsequently, to utilize the captured spatial and semantic relationships, we employ two separate, parallel strands of graph convolution modules to extract the aforementioned fuller global information. The graph convolution modules are based heavily on the graph convolutional networks introduced by Kipf et al. [37], resembling two two-layered simple graph convolutional networks employed side-by-side, one utilizing the spatial adjacency matrix and theother the semantic adjacency matrix. Specifically:

$$G^{sp} \in \{V, E\}, V = F, E = A^{sp} \tag{5}$$

$$\begin{cases} h_i^{sp} = \sigma(A^{sp} h_{i-1} w_i), \ i \in \{1, 2\} \\ h_i^{sp} = V, \ i = 0 \end{cases} \tag{6}$$

$$G^{sp} \in \{V, E\}, V = F, E = A^{se} \tag{7}$$

$$\begin{cases} h_i^{se} = \sigma(A^{se} h_{i-1} w_i), \ i \in \{1, 2\} \\ h_i^{se} = V, \ i = 0 \end{cases} \tag{8}$$

Where h_i represent the input/output i-th layer of the graph convolution modules, and w_i represent the learnable parameters.

To finally combine the two distinct types of information extracted (spatial and semantic), we employ the same merge function used to combine the outputs of the encoder's third and fourth layer for the graph convolutional modules to return an enhanced global representation of the features as their combined output:

$$F_d = \text{conv}((h_2^{sp} \parallel h_2^{se}), \theta_{F_d}) \tag{9}$$

Figure 3 shows the detailed illustration of the global semantic context extractor

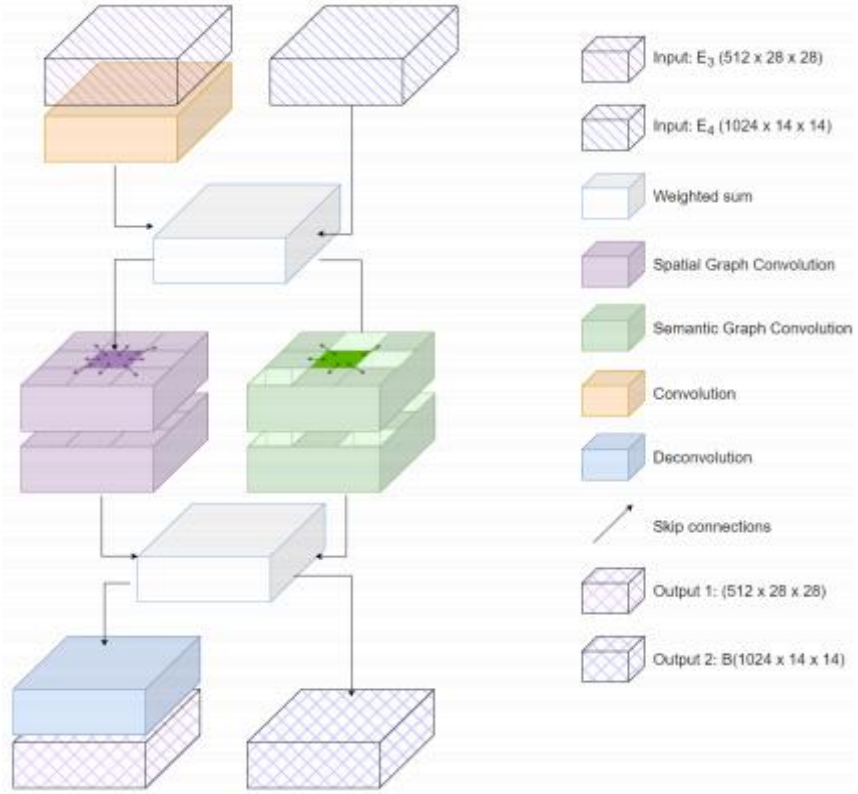

Figure 3. An detailed illustration of the global semantic context extractor

3.4. Final Decoder

To utilize the enhanced global representation, we then input the result into the decoder of the U-Net backbone, which we have set to adopt a similar approach as the initial encoder. Specifically, the previous output is inputted as a skip connection to the two bottom-most layers of the decoder (the third and the fourth, corresponding with the encoder), and is computed through ascending decoder blocks with direct skip connections from the encoder before arriving at our final prediction:

$$\begin{cases} D_i = \mathrm{up}(D_{i-1}, E_i), \ i \in \{2,1,0\} \\ D_i = \mathrm{up}(D_{i-1}, \mathrm{deconv}(F_d)), \ i = 3 \\ D_i = \mathrm{up}(E_4, F_d), \ i = 4 \end{cases} \tag{10}$$

$$P_n = D_0 \tag{11}$$

Here, P_n denotes the final prediction for data set n, and up $(\cdot)$ denotes the decoder blocks, whose configurations are similarly listed in fig. 2.

3.5. Optimization

For optimization, we utilize a joint-loss function combining dice loss and binary-cross-entropy loss, allowing for enhanced stability and convergence and improved handling of class imbalance and boundary detection. Specifically:

Where:

$$\mathcal{L}(P_n, Y_n) = \alpha \mathcal{L}_{dice}(P_n, Y_n) + \beta \mathcal{L}_{bce}(P_n, Y_n) \tag{12}$$

$$\mathcal{L}_{dice}(p, y) = -\sum_{i=0}\sum_{i=0} y_{i,j} \cdot log(P_{i,j}) \tag{13}$$

$$\mathcal{L}_{bce}(p, y) = \frac{\sum_{i=0}^{H-1}\sum_{i=0}^{W-1} p_{i,j} y_{i,j}}{\sum_{i=0}^{H-1}\sum_{i=0}^{W-1} {p_{i,j}}^2 + \sum_{i=0}^{H-1}\sum_{i=0}^{W-1} {y_{i,j}}^2} \tag{14}$$

4. Experiments

4.1. Dataset Information

In order to test the result of our GUNet model, we used 3 datasets to conduct various experiments and evaluated the overall performance. The datasets that are used in the evaluation are CVC-ColonDB, ETIS and Kvasir. The three different datasets all contain images of colonoscopy exams images. The CVC-ColonDB dataset contains 379 training sets of original resolution 574 ×500, the ETIS data set contains 196 images of initial resolution1225 × 966, and the Kvasir dataset contains 1000 images of varying resolutions. While all the 3 datasets contain training sets that have different resolutions, we resized all of the images that are used in the training process of the model to the resolution of 244 × 244 × 3. The images contained in these datasets are real images that were taken on standard colonoscopy exams. The images show the abnormal tissue walls of the colon, potentially being suspected of signs showing colorectal cancer that were noticed by doctors. To complete the experiment and increase the complexity of the images that were being analyzed by the model, we not only included the ones that clearly show unusual changes on the tissue walls but also the ones that are notclear and hard for humans to identify problematic tissue walls.

4.2. Evaluation Criteria

To evaluate the performance of GUNet and the other models used for comparison, we utilize five commonly used evaluation metrics. They are the Jaccard Index (JA), Dice coefficient (DI), Sensitivity (SE), Specificity (SP), and Accuracy (AC). Specifically, with TP, TN, FP, FN denoting true positives, true negatives, false positives, and false negatives respectively, the Jaccard Index formula used to calculate metrics can be defined as:

$$JA = \frac{TP}{TP + FN + FP} \tag{15}$$

The Dice coefficient formula can be defined as:

$$DI = \frac{2TP}{2TP + FN + FP} \tag{16}$$

The Sensitivity formula can be defined as:

$$SE = \frac{TP}{TP + FN} \tag{17}$$

The Specificity formula can be defined as:

$$SP = \frac{TN}{TN + FP} \tag{18}$$

The Accuracy formula can be defined as:

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

4.3. Results

Table 1. A comparison of GUNet with other approaches on the combined dataset

method	JA(%)	DI(%)	SE(%)	SP(%)	AC(%)
UNet [11]	64.23	71.16	71.81	97.13	94.23
GUNet	**64.47**	**71.26**	**72.67**	**97.38**	**94.67**

In order to test the effectiveness of our proposed GUNet, we ran several other models including UNet and evaluated the performance using the same evaluation criteria - Jaccard Index (JA), Dice Coefficient (DI), Sensitivity (SE), Specificity (SP) and Accuracy (AC). Thedata showing the percentage of accuracy of all the evaluation criteria for GUnet and UNet is shown in table 1. According to our data, we can see that all five evaluation criteria show that our proposed GUNet performs better than the traditionally used UNet in terms of accuracy for segmenting the abnormal parts of the colorectal exam images. In terms of the Jaccard Index, the accuracy of UNet on segmentation is around 64.23%, while our GUNet model obtained an accuracy of 64.27%, with 0.04% better improvements. Moving on to the Dice Co efficient, the UNet has an accuracy of 71.16%, in which GUNet achieved a better result, 0.1% higher than UNet, with 71.26% accuracy. Next, in terms of the evaluation done with Sensitivity, while the UNet only achieved an accuracy of 71.81%, our model is performing much better, with 72.67% accuracy. When evaluating performance using Specificity criteria, the accuracy of UNet turns out to be 97.13%, while our GUNet got a result of 97.38%, more accurate than UNet. Finally, the accuracy of UNet when evaluating using Accuracy evaluation criteria is 94.23%, whereas our proposed GUNet got and higher result of 94.67%. To sum up, by examining the data in table 1, we can conclude that our proposed GUNet model is performing better than the traditional UNet in terms of accuracy in terms of segmentation. This indicated that using GUNet instead of traditional method will be able to improve the accuracy of segmentation of images of colorectal exams in order to identify unusual parts of the colon, and eventually benefit the patients by more accurately and preciously finding early signs of colorectal cancer.

5. Experiments

In this paper, we propose a novel encoder-decoder architecture for colorectal polyp segmentation. In order to obtain fuller global representations and extract enhanced global information, we came up with a novel graph convolution module, which calculates and combines both spatial and semantic global relationships in parallel within the feature maps. This allows for improved contextual information with the merging of spatial reasoning and semantic understanding. In addition, we complementarily combine feature maps of different depths as input for the module in order to achieve a fuller representation of semantic information, and we utilize a combination of three individual datasets for a greater quantity of data and generalization of the model. Experimentation demonstrates that GUNet can serve as a competitive alternative to the U-Net, surpassing it when training on the combined dataset.

Acknowledgements

We would like to express our appreciation firstly for our parents, for sup-porting us in our research; secondly Mr. Jordan King, for his considerable contributions in the conceptualization phase; and thirdly for the opportunity to be able to participate in the Yau High School Science Award, for the platform that we got to showcase our knowledge in programming. Through the process, we not only got a chance to research more into Colorectal Cancer but also to experiment with our own approach to provide a better solution to the identification of early signs of cancer.

References

[1] Rebecca L. Siegel et al. “Colorectal cancer statistics, 2023”. In: CA: A Cancer Journal for Clinicians 73.3 (2023), pp. 233 – 254. issn: 1542 - 4863. doi: 10.3322/caac.21772.

[2] Lou Guo-chun et al. “A retrospective study on endoscopic missing diagnosis of colorectal polyp and its related factors”. In: The Turkish Journal of Gastroenterology 25.1 (Apr. 22, 2015), pp. 182 – 186. issn: 13004948, 21485607. doi: 10.5152/tjg.2014.4664.

[3] A. Leufkens et al. “Factors influencing the miss rate of polyps in a back to-back colonoscopy study”. In: Endoscopy 44.5 (May 2012), pp. 470 – 475. issn: 0013 - 726X, 1438-8812. doi: 10.1055/s-0031-1291666.

[4] V Prasath. “Polyp Detection and Segmentation from Video Capsule Endoscopy: A Review”. In: Journal of Imaging 3.1 (Mar. 2017). Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, p. 1. issn: 2313 - 433X. doi: 10.3390/jimaging3010001.

[5] Stavros A. Karkaniset al. “Computer-aided tumor detection in endoscopic video using color wavelet features”. In: IEEE transactions on information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology Society 7.3 (Sept. 2003), pp. 141 – 152. issn: 1089 - 7771. doi: 10.1109/titb.2003.813794.

[6] Lu´ıs Alexandre, Nuno Nobre, and Jo˜ao Casteleiro. “Color and Position versus Texture Features for Endoscopic Polyp Detection”. In: 2008 International Conference on BioMedical Engineering and Informatics. 2008 International Conference on BioMedical Engineering and Informatics. Vol. 2. ISSN: 1948 - 2922. May 2008, pp. 38 – 42. doi: 10.1109/BMEI.2008.246.

[7] Alain S´anchez-Gonz´alez et al. “Automatized colon polyp segmentation via contour region analysis”. In: Computers in biology and medicine 100 (Jan. 9, 2018). Publisher: Comput Biol Med. issn: 1879-0534. doi: 10. 1016/j.compbiomed.2018.07.002.

[8] Geert Litjenset al. “A survey on deep learning in medical image analysis”. In: Medical Image Analysis 42 (Dec. 2017), pp. 60 – 88. issn: 13618415. doi: 10.1016/j.media.2017.07.005.

[9] Syed Muhammad Anwar et al. “Medical Image Analysis using Convolutional Neural Networks: A Review”. In: Journal of Medical Systems 42.11 (Nov. 2018), p. 226. issn: 0148-5598, 1573-689X. doi: 10.1007/s10916 - 018 - 1088 - 1.

[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. Mar. 8, 2015. arXiv: 1411.4038 [cs].

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. May 18, 2015. arXiv: 1505.04597 [cs].

[12] Debesh Jha et al. ResUNet++: An Advanced Architecture for Medical Image Segmentation. Nov. 16, 2019. arXiv: 1911.07067 [cs, eess].

[13] Debesh Jha et al. “Real-Time Polyp Detection, Localization and Segmentation in Colonoscopy Using Deep Learning”. In: IEEE Access 9 (2021), pp. 40496 – 40510. issn: 2169-3536. doi: 10.1109/ACCESS.2021.3063716.

[14] Zongwei Zhou et al. “UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation”. In: IEEE Transactions on Medical Imaging 39.6 (June 2020), pp. 1856–1867. issn: 0278 - 0062, 1558 - 254X. doi: 10.1109/TMI.2019.2959609.

[15] Huimin Huang et al. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. Apr. 19, 2020. arXiv: 2004.08790 [cs, eess].

[16] Tanvir Mahmud, Bishmoy Paul, and Shaikh Anowarul Fattah. “PolypSeg-Net: A modified encoder-decoder architecture for automated polyp seg-mentation from colonoscopy images”. In: Computers in Biology and Medicine 128 (Jan. 2021), p. 104119. issn: 00104825. doi: 10.1016/j.compbiomed.2020.104119.

[17] Ozan Oktay et al. Attention U-Net: Learning Where to Look for the Pancreas. May 20,2018. arXiv: 1804.03999 [cs].

[18] Zijin Yin et al. Duplex Contextual Relation Network for Polyp Segmentation. Jan. 19, 2022.arXiv: 2103.06725 [cs, eess].

[19] Francesco Visinetal. ReSeg: A Recurrent Neural Network-based Model for Semantic Segmentation. May 24, 2016. arXiv: 1511.07053 [cs].

[20] Mahmood Haithami et al. “An Embedded Recurrent Neural Network based Model for Endoscopic Semantic Segmentation”. In: (2021).

[21] Sae Hwang et al. “Polyp Detection in Colonoscopy Video using Elliptical Shape Feature”. In: 2007 IEEE International Conference on Image Processing. 2007 IEEE International Conference on Image Processing. San Antonio, TX, USA: IEEE, 2007, pp. II – 465 – II – 468. isbn: 978-1-4244 – 1436 - 9. doi: 10.1109/ICIP.2007.4379193.

[22] C. Van Wijk et al. “Detection and Segmentation of Colonic Polyps on Implicit Isosurfaces by Second Principal Curvature Flow”. In: IEEE Transactions on Medical Imaging 29.3 (Mar. 2010), pp. 688–698. issn: 0278 - 0062, 1558 - 254X. doi: 10.1109/TMI.2009.2031323.

[23] Juan Silva et al. “Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer”. In: International Journal of Computer Assisted Radiology and Surgery 9.2 (Mar. 2014), pp. 283 – 293. issn: 1861-6410, 1861-6429. doi: 10.1007/s11548 - 013 - 0926 - 3.

[24] Alexander V. Mamonovetal. “Automated Polyp Detection in Colon Capsule Endoscopy”. In: IEEE Transactions on Medical Imaging 33.7 (July 2014), pp. 1488–1502. issn: 0278-0062, 1558-254X. doi: 10.1109/TMI. 2014.2314959.

[25] Yuji Iwahori et al. “Automatic Detection of Polyp Using Hessian Filter and HOG Features”. In:Procedia Computer Science 60 (2015), pp. 730 – 739. issn: 18770509. doi: 10.1016/j.procs.2015.08.226.

[26] Jorge Bernal et al. “WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validationvs. saliency maps from physicians”. In: Computerized Medical Imaging and Graphics 43 (July 2015), pp. 99 – 111. issn: 08956111. doi: 10.1016/j.compmedimag.2015.02.007.

[27] Foivos I. Diakogiannis et al. “ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data”. In: ISPRS Journal of Photogrammetry and Remote Sensing 162 (Apr. 2020), pp. 94–114. issn: 09242716. doi: 10.1016/j. isprsjprs.2020.01.013. arXiv: 1904. 00592 [cs].

[28] Jiahuan Song et al. “Global and Local Feature Reconstruction for Medical Image Segmentation”. In: IEEE Transactions on Medical Imaging 41.9 (Sept. 2022), pp. 2273–2284. issn: 0278-0062, 1558-254X. doi: 10.1109/ TMI.2022.3162111.

[29] Debapriya Banik et al. “Polyp-Net: A Multimodel Fusion Network for Polyp Segmentation”. In: IEEE Transactions on Instrumentation and Measurement 70 (2021), pp. 1–12. issn: 0018-9456, 1557 - 9662. doi: 10. 1109/TIM.2020.3015607.

[30] Xinzi Sun et al. Colorectal Polyp Segmentation by U-Net with Dilation Convolution. Dec. 26, 2019. arXiv: 1912.11947 [cs, eess].

[31] Krushi Patel, Andres M. Bur, and Guanghui Wang. “Enhanced U-Net: A Feature Enhancement Network for Polyp Segmentation”. In: 2021 18th Conference on Robots and Vision (CRV). 2021 18th Conference on Robots and Vision (CRV). Burnaby, BC, Canada: IEEE, May 2021, pp. 181 – 188. isbn: 978 - 1 - 66541 - 413 - 5. doi: 10.1109/CRV52889.2021.00032.

[32] Xiaoqi Zhao, Lihe Zhang, and Huchuan Lu. Automatic Polyp Segmentation via Multi-scaleSubtraction Network. Aug. 11, 2021. arXiv: 2108. 05082 [cs].

[33] Abhishek Srivastava et al. GMSRF-Net: An improved generalizability with global multi-scale residual fusion network for polyp segmentation. Nov. 20, 2021. arXiv: 2111.10614 [cs, eess].

[34] Guanghui Yue et al. “Boundary Constraint Network with Cross Layer Feature Integration for Polyp Segmentation”. In: IEEE Journal of Biomedical and Health Informatics 26.8 (Aug. 2022), pp. 4090 – 4099. issn: 2168 - 2194, 2168 - 2208. doi: 10.1109/JBHI.2022.3173948.

[35] Guanghui Yue et al. “Attention-Guided Pyramid Context Network for Polyp Segmentation in Colonoscopy Images”. In: IEEE Transactions on Instrumentation and Measurement 72 (2023), pp. 1 – 13. issn: 0018 - 9456, 1557 - 9662. doi: 10.1109/TIM.2023.3244219.

[36] Ruifei Zhang et al. Adaptive Context Selection for Polyp Segmentation. Jan. 11, 2023. arXiv: 2301.04799 [cs].

[37] Thomas N. Kipf and Max Welling. “Semi-Supervised Classification with Graph Convolutional Networks”. In: (Feb. 22, 2017). doi: 10.48550/ arXiv.1609.02907.