NVEO
Natural Volatiles & Essential Oils

SU-NET Based colorectal polyp Segmentation from Colon Cancer Morphology Images

Mohan Mahanty[1,*] Debnath Bhattacharyya [2,*], Divya Midhunchakkaravarthy[1]

[11]Department of Computer Science and Multimedia, Lincoln University College, Malaysia.

[2,*]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Field, Vaddeswaram, Guntur 522502, India.

*Corresponding author. Email: debnathb@kluniversity.in

Abstract

Precise demarcation of glands from clinical histology images are pre-requirement for accurate medical diagnosis. Colorectal polyps that originate and expandsover the rectum or colon membrane are the decisive reason for colorectal Cancer(CRC). The early-stage recognition and of polyps and treatment can decrease the mortality rate. To lower the polyp miss-rate in colonoscopy, a Computer-Aided Medical Diagnosing(CAD) system with high accuracy is needed. In recent times, researchers develop deep learning models for accurate polyp detection from histomorphology images, but accuracy is still the most requisite factor for reliable results. In this paper, we propose to develop and test a Convolutional Neural Network(CNN) based U-shape network (SU-NET) model for semantic segmentation of colorectal polyps from colonoscopy images.SU-NET is an Encoder-Decoder-based architecture, inspired by the popular segmentation architectures SegNet and U-Net for improved colon polyp segmentation. In the proposed model the top most layers transfer the Pooling indices whereas the lower-level layers transfer the feature-maps to incorporate fine multiscale information for better colon polyp contour identification.We evaluated the proposed algorithm in contrast with various prominentdeep learning architectures across multi-modal biomedical image segmentation tasks to segment polyps from the colonoscopy and histopathology images.For evaluating the proposed model, an accredited and publicly available colonoscopy image dataset CVC-ColonDB is employed. The model achieves a recall of 91.3%, F_1-Score of 90.81%, F_2-Score of 86.39%, Precision of 89.21%, and the Dice similarity coefficient of 0.895 outshines the existing advanced deep learning CNN models.

Keywords: Colorectal cancer, Colonoscopy, Polyp semantic segmentation, medical image analysis, deep convolutional neural network, SU-NET

1. Introduction

Colon or rectal cancer occurs either in the colon or in the anus (rectum). Colon cancers most often begin as a colorectal polyp, which originates inside the colon or anus that may, later on, turn out as the most crucial reasons for the intestinal cancer cells. Digestive tract polyps are the reason for the development of 80-90% of the intestinal cancers. In 2020, around 147,950 individuals are diagnosed with colorectal cancer, and the mortality rate is supposed to 53,200. Alaska natives are highly suffering from CRC, whereas the intensity of risk is lower for the Asia pacific islanders. And the study states that 30 to 40 % of men are more affected than women [1]. In recent years, the frequency of detection and death rate of colon cancer has raised up.The colon is the longest part of the human body, consisting of millions of glands, so the traditional methods cannot detect the tiny polyps that may lead to a malignant tumor in the future. All polyps may not lead to cancer, but polyp detection plays a significant role in decreasing the human mortality rate. Colon cancer cells usually develop as a polyp on the intestinal mucosa's inward lining. However, it can additionally exist as an origin for a benign lesion called an adenoma that can change into a malignant lesion relying on its histological discussion as received in table 1.

Table 1: Histological variety and malignancy association

Histological Type	Percentage of Colorectal cancer cases	Percentage of Malignancy
Tubular	60-80%	<5%
Tubulovillous	10-25%	20-25%
Villous	5-10%	35-45%

The people with obesity and high alcoholic consumption are mostly affected with CRC but still the accurate causes are not identified. Having nicotine addiction and desk-bound habits has more chance of causing colon cancer[2]. People whose age is greater than 50 are usually being diagnosed by this disease, but the recent statistics representing that it is more aggressive at younger ages. The pathologists used an effective prognosis evaluation technique like colonoscopy which helps in screening and medical diagnosis of intestinal cancer cells.

Based on the eye-hand synchronization and operating skills of the gastroenterologist, the diagnosis method got succeeded. This diagnosis procedure is lengthy and expensive, and the discarding of benignant polyps for biopsy symbolizes the system's inadequacy. The identification of polyps in the early stages is difficult because of their minute sizes and shapes. Analysis of polyps conducted by the pathologists is a crucial task and time-consuming traditional process. It is highly unreliable as the classification and grading of colon polyps need considerable data regarding their specific form, size, structure, and general inheritor morphological features.

Computer-assisted medical images Diagnosis depends upon highly intricate pattern recognition, offered for automatic detection of the dubious colon or rectal polyps in the colon CT images. By developing Computer Assisted Pathological Diagnosis (CAPD), Deep learning architectures, the gastroenterologist can quickly diagnose the adenocarcinomas at different magnifications from healthy tissues of the digitized colon histological images. Researchers are developing computer-aided trustworthy systems for malignant polyp detection[3] to overcome the complication in conventional microscopic colon cancer detection techniques. Deep learning architectures evolution changed the researchers motive towards cancer research because they can detect the nano millimeter-sized malignant tissues (polyps). Therefore, this paper was highly encouraged by the need to get an early and precise segmentation of a polyp from the clinical histopathology images.

2. Related work

Farah Deeba et al. proposed a colon polyp detection method based on image enhancement and saliency-based selection[4]. They applied their model to colonoscopy and WCE images and evaluated on openly available colon databases. They claimed that their model achieved a recall of 86.33% and an F2 score of 75.51% for the CVC Colon DB[5] dataset. Automated Region-based Localization of Polyps was proposed by Sudhir Sornapudi et al. [6] presented an enhanced region-based CNN by creating masks around the polyps. They evaluated their proposed work on openly available colon databases and attained an F1 score of 90.73 and an F2 score of 91.27 on the CVC-Colon DB Dataset.

Ashkan Tashk et al.[7] proposed Automatic Segmentation of Colorectal Polyps using a Novel and Innovative CNN. They make use of the modified U-net architecture for sophisticated semantic segmentation of polyps from the optical colonoscopy(OC) images. They evaluated the proposed CNN model with the well-known OC image databases and claimed that their model attains a Precision of 62.0, Recall 82.4, F1- Score 70.7 on CVC-ColonDB Dataset.

Jaeyong Kang et al. proposed their work on Instance Division Designs for Polyp Division in Colonoscopy Images[8]. They ensembled ResNet 50 and ResNet 101 based Mask R-CNN models to improve the segmentation efficiency. They evaluated their model on openly available colon databases. The authors claimed that their model achieves a mean-Precision of 77.92, mean pixel recall of 76.25, and IU

(interception over union) of 69.46 on the CVC-Colon DB dataset.

Ngoc-Quang Nguyen et al. suggested a Consecutive Deep Encoder-Decoder network for robust contour segmentation in the pathology images[9]. They described that their model could hold the multi-level contextual information and learn abundant details of missed pixels features at the training phase. And it can also capture the contours of the objects by using effective multiscale decoders. They claimed that their CDED-net model attained an accuracy of 0.980, Specificity of 0.991, Dice score of 0.896, Level of sensitivity of 0.792 on the CVC-Colon DB dataset.

DebapriyaBanik et al.[9]proposedan enhanced version of DT-WpCNN named as Polyp-Net, used the LG-LSM to reduce the false-positive rate. The model is trained and tested over the CVC-colon DB dataset and achieves a dice index of 0.839, Precision of 0.836,volume-similarity of 0.863, F1-score of 0.823, Recall of 0.811, Hausdorff distance of 21.796, and F2-score of 0.815, which outperforms the existing state-of-the-art methods.

Le Thi Thu Hong et al. proposed an Ensemble of U-Nets with Efficient Net to segmentation polyps from the colonoscopy images[11]. They used the transfer learning method and an ensemble of 2 U-net models with two variations of Efficient Net for accurate segmentation of colon polyps. They also adapted the asymmetric similarity loss function for better results in between Precision and Recall. The model is evaluated on the well-known colon polyp datasets and attains 89.13% of dice Score, 79.77% IOU, 90.15% recall, and 86.28% precision on the CVC-Colon DB Dataset.

Hemin Ali Qadir et al. [12] proposed the Single-shot feed-forward FCNN for real-time polyp detection and segmentation. They used the 2D Gaussian masks for accurate detection of different sized, shaped polyps effectively. They evaluated their model on the CVC-Colon DB, ETIS-LARIB datasets. After evaluation, they claimed a recall of 91%, a precision of 88.35%, and an F1 score of 89.65% over the CVC-Colon DB, which are considered good compared to other models. Deep learning models proposed by the researchers shows a significant impact on the medical pathological image analysis. Colorectal polyp segmentation plays an imperative role in diagnosing colorectal cancer. Meticulous colon polyp detection is still a complex problem in computer-aided medical diagnosis due to the fuzzy boundaries and non-differentiable complex backgrounds. To overcome this, we propose a SU-NET deep learning model for accurate segmentation of the colorectal polyps.

3. Proposed Architecture for gland segmentation

Computer-assisted medical image analysis pathology studies the causes and pathogenesis of disease occurrence. Deep neural networks in computer-based medical image diagnosis precisely identify abnormal gland images and help pathologists choose reliable treatment. Accurate detection of the polyps through image segmentation is still a challenge in medical diagnosis. This paper proposed a unique deep learning model for meticulous colorectal polyp detection from the colonoscopy images. Figure 1, represents the architecture of the proposed model.

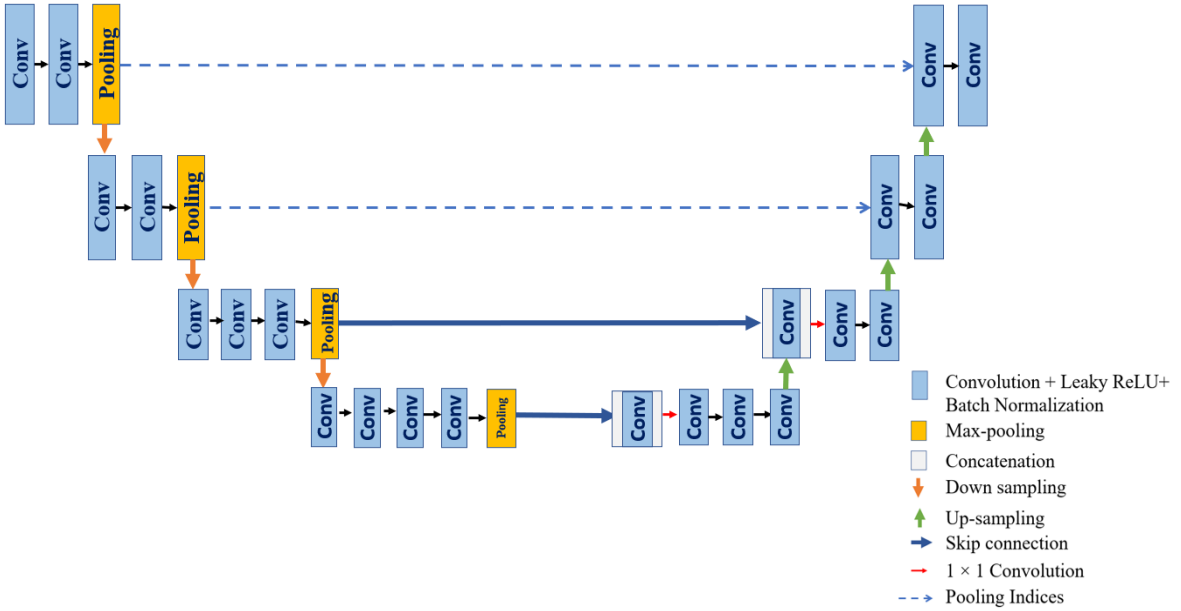


Fig 1. Proposed SU-NET architecture

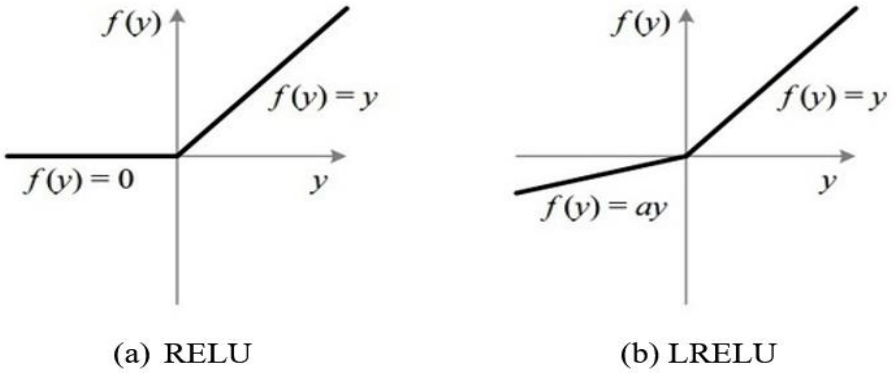
The proposed architecture doesn't consist of a fully connected (FCNN) layer at the bottleneck, so the architecture's memory and complexity are reduced. The SU-NET architecture includes an encoder and an appropriate decoder network, followed by a pixel-wise classification layer (Softmax). The proposed encoder network contains 11 convolutional layers, considered from the first 11 layers of the VGG16 network[13]. And the corresponding Decoder network also consists of 11 de-convolutional layers.

3.1 Encoder

The contracting path on the left-hand side consists of 4 blocks. The first and second blocks consist of two convolutional (conv) layers. The third block consists of three conv layers, and the fourth block consists of four conv layers. Each block of the encoder network consists of a series of convolutional layers, which perform numerous convolutional operations with the provided kernels (size 3 × 3, by stride 1) and generates a set of feature maps which are then Group Normalized(GN)[14].

GN divides the total channels into small groups and computes the mean and variance within each group for normalization. GN is applied on the batches independently and attains a stable accuracy than Batch normalization (BN)[15] on various batch sizes. When the batch size is small, the number of errors increases with BN due to the inaccurate batch statistics. Then an element-wise- leaky ReLU[16] is applied, and following that, max-pooling is applied (2×2 size, stride 2), then the result generated is sub-sampled by factor 2.

All the existing segmentation models used the RELU, where ReLU obtained the results from max (0, x). When it takes the negative slope values as input, it completely blocks the learning in the ReLU because of gradients of 0 in the negative part, and it sets the output to zero. While training a deep neural network, if a neuron gets negative during the back propagation stage, it can’t recover back, i.e.,they do not have a converge to a good local minimum, considered as a "dead neuron. Furthermore, these dead neurons never played any role in the neural network and diminished the network's performance, known as the "dying ReLU problem". Figure 2 graphically describes the working of ReLU and LReLU.



(a) RELU (b) LRELU

Fig 2. Graphical representation of (a) RELU (b) LRELU

ReLU set the output value to zero(0) when the neuron met with a negative value. As represented in Eq(1) and Eq(2), Leaky ReLU overcomes the dying ReLU problem by setting some slope (small value say 0.001 (α)) for negative values instead of a flat slope.

$$f(x) = \max(0.001x, x) \quad \text{Eq (1)}$$

$$f(x)=1(x<0)(\alpha x)+1(x>=0) \quad \text{Where } \alpha=0.001 \text{ is constant.} \quad \text{Eq (2)}$$

Where αx will have a non-zero value, and it will continue learning without reaching a dead end. Hence, leaky ReLU performs better than ReLU.

For each down-sampling, to extract the input feature maps, we need to double the number of kernels used for convolutional operations. A series of max-pooling and sub-sampling operations reduces the spatial resolution of the feature maps, which affects the segmentation process of the objects in the input image. So, it's necessary to store the encoded feature maps after max pooling. Pooling indices generated at the uppermost two layers of the encoder network are transferred to the corresponding layers of the decoder network. And the high-resolution feature maps generated at the lowermost layers of the encoder network,transferred to the appropriate decoder network, and combined with the up-sampling output.

3.2 Decoder

The decoder network consists of 4 blocks. The lower-level blocks consist of four de-convolutional layers. The next upper block consists of three deconvolutional layers, and the subsequent two uppermost blocks consist of two de-convolutional layers. The two lowermost blocks of the decoder network perform the de-convolution (up-convolution) using the kernels and appropriate feature maps received from the corresponding encoder blocks (skip connections inspired from the U-net architecture)[17]. And then generates the feature maps, which thengroup normalized[14]. After convolutions biases are not included and there is no application of leaky ReLU is not applied on the decoder network.

For each up-sampling, the number of kernels used for convolutional operations is halved, used for deconvolutional operations to generate the object's non-lossy boundary information. In the uppermost two blocks of the decoder network, inverse convolution is performed, using a trainable decoder of this multi-channel up-sampling kernel. In up-sampling, fixed bicubic interpolation weights are used because no learning is there in the up-sampling process and the max-pooling indices received from the encoder network are used to densify its sparse input. We used the bicubic interpolation instead of bilinear interpolation, which is used in SegNet[18].

$$v(x,y) = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij} x^{i} y^{j} \quad \text{Eq(3)}$$

.

As shown in Equation (3), Bilinear- interpolation considers the average of those block of four pixels (2x2), yields to blur edges. Whereas Bicubic-interpolation considers the nearest 16 pixels (4x4) for each missed pixel, yields smoother and accurate edges. Finally, the output generated from the last decoder block is given as input to a multi-class classifier layer (Softmax).

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad \text{Eq(4)}$$

As shown in Equation 4, the output generated from the Softmax is a K-channel image, where k is the total number of classes of objects in the image.

4. Experimental Results and Analysis

4.1 Dataset

We consider the CVC ColonDB[5] colon polyp dataset, which consists of 380 images extracted from 15 short different colonoscopy video studies, for several examples of polyps. As shown in figure 3, each image consists of only one polyp and annotated by gastroenterologists. The database comprises original images, Polyp masks, and Polyp contours. All the images have a fixed-size resolution of 574x500. Clinical experts provide binary masks for every image of the dataset. The Ground Truth(GT) masks allude to the exact boundaries around the polyp regions shown in white and black.

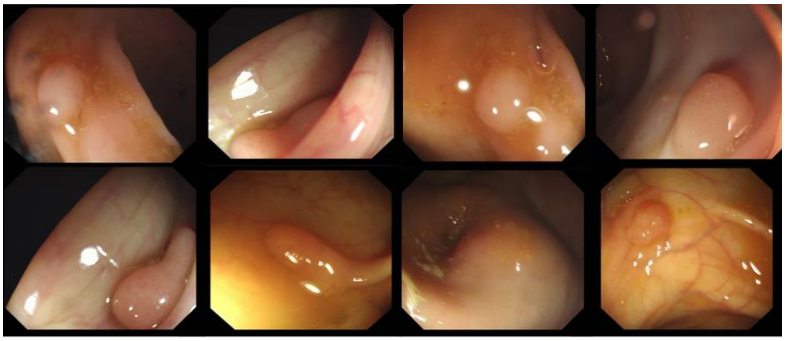


Fig.3. Appearance variability of polyps

4.2 Data Augmentation

Data augmentation is a crucial strategy commonly used in deep learning. The major obstacle in clinical image analysis is the availability of image data. Due to personal privacy reasons and lack of coordination between pathologists and researchers, the required image data may not be available during the model training, which results in declination of the model's performance. Depend upon the pathologists and equipment used in colonoscopy, different orientations of polyp images are generated. The data augmentation technique brings colonoscopy images right into an extensive area that can cover all their variations, increase the robustness, and lower our proposed model's over-fitting.

The CVC Colon DB dataset, primarily composed of 380 images, has a fixed size resolution of 574× 500. After removing the canvas around the original image, the obtained informative image is resized to 512 × 512 before training to be suitable for processing on the proposed SU-NET. To enhance the number of images in training data, image-level data augmentation is applied. We considered general geometric transformations in the augmentation process. We used the rotation operation in the augmentation process, where each image in the original dataset is rotated by 90°,180°, and -90°.Then on the obtained images, vertical and

horizontal flipping is applied to generate a total of 3040 images.We followed the same process for the augmentation of corresponding Ground Truth images.

4.3 Model Evaluation

The proposed model was trained and tested on different data sets known as holdout evaluation, which provides an unbiased estimate of learning performance. The general criteria to clinically evaluate the proposed computer-aided diagnosis (CAD) SU-NET model for polyp segmentation, implemented on standard CVC-Colon DB polyp database images, we use the metrics, sensitivity(Recall), F1 Score, F2 Score to compare our Segmented polyp (SP) with the ground truth (GT) images. Based on the confusion matrix, we get four types of cases,

True positive (TP): Polyp pixels segmented correctly.

True negative (TN): Non-polyp pixels predicted as non-polyp pixels.

False positive (FP): Background pixels classified as polyp pixels.

False negative (FN): Polyp pixels identified as non-polyp pixels.

Sensitivity: Sensitivity (Recall) measures the ratio of the positive values considered by the segmentation process and the correct positive values given by the ground truth.Recall is defined as in equation 5.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad \text{Eq(5)}$$

Precision: The ratio of true positive outputs to the total number of predicted results, including false positives. The total predicted positive instances can be represented using equation 6 as,

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad \text{Eq (6)}$$

The F_β-measure is calculated to offer a basic sight of the efficiency of our method. It is a generalization of the F-measure, calculated as the harmonic mean of the Precision and sensitivity, of the same weight. Good F-Score implies that you have a smaller number of FP and FN. Equation 7 describes the mathematical notation of F_β-Measure.

$$F_\beta = \frac{(1+\beta^2)*\text{Precision}*\text{Recall}}{\beta^2*\text{Precision}+\text{Recall}} \quad \text{Eq (7)}$$

When Precision and recalls have the same weight, F_1-Score is calculated by considering the β=1, as shown in equation 8.

$$F_1\text{-Score} = \frac{(1+(1)^2)*\text{Precision}*\text{Recall}}{(1)^2*\text{Precision}+\text{Recall}} \quad \text{When } \beta=1 \quad \text{Eq(8)}$$

$$= \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad \text{Eq(9)}$$

After simplification, we get Equation 9, and if we substitute the Precision and recall with their equations, we get the F_1-Score in terms of True Positives and Negatives as in the form of equation 10.

$$F_1\text{-Score} = \frac{2*TP}{2*TP+FP+FN} \quad \text{Eq(10)}$$

When there is a low weight on Precision and more weight to recall, we get the F_2-Score, as shown in equation 11.

$$F_2\text{-Score} = \frac{(1+(2)^2)*Precision*Recall}{(2)^2*Precision+Recall} \quad \text{When } \beta=2 \quad \text{Eq(11)}$$

$$= \frac{5*Precision*Recall}{4*Precision+Recall} \quad \text{Eq(12)}$$

After simplification, we get Equation (12), and if we substitute the Precision and recall with their equations, we get the F_2-Score in terms of True Positives and Negatives as in equation 13.

$$F_2\text{-Score} = \frac{5*TP}{5*TP+4*FN+FP} \quad \text{Eq(13)}$$

We also considered the Dice similarity coefficient as a segmentation evaluation metric to compare the similarity between prediction images(S) and the Ground Truth(G) images.

$$Dice(G,S) = 2*\frac{|G \cap S|}{|G|+|S|} \quad \text{Eq(14)}$$

The mathematical notations shown in equation 14, G and S determine ground truth and segmented binary labels.

4.4 Experimental Setup

All the experiments were performed on a computer with a 2.4 GHz Intel(R) i7-7th gen processor, CPU with 32 GB of RAM, and NVIDIA GeForce Titan X GPU. The model execution has been conducted with Python 2.7 on Keras and TensorFlow[19]. Since we don't use a pre-trained model, we have to augment our training data. The proposed architecture is trained by using the 3040 augmented images is conducted over 80 epochs.

4.5 Comparative Results Analysis

After Establishing the SU-NET Model, it is evaluated on the benchmark CVC-ColonDB dataset as a concluding step. We divided the generated augment dataset (3040 images) into a training dataset (2432 images, which is 80 % of 3040 images) and a testing dataset (608 images, which is 20% of 3040 images). The proposed SU-NET is a deeper CNN model with a considerable number of learnable parameters, so it takes a little more time for training. Still, can extract the feature from the images accurately. The results depicted in Table2 described the proposed model's segmentation results evaluated using the metrics Recall, F_1- Score, and F_2- Score.

Table 2. Comparative Segmentation results w.r.t Recall, F_1-Score and F_2-Score

Model	Recall	F_1-Score	F_2-Score
Farah Deeba et al. [4]	86.33	-	75.51
Sudhir Sornapudi et al. [6]	-	90.73	91.27
Ashkan Tashk et al.[7]	82.4	70.7	-
Hemin Ali Qadir et al. [12]	91.00	89.65	-
Proposed SU-NET	91.32	90.81	86.39

The F-measure is considered a combined metric. It measures a test's accuracy using Precision and Recall, which is considered perfect when it's 1 and 0 while the model is completely failed.The quantitative results of Table 2 demonstrate that our proposed model obtained the Recall of 91.3 % (Recall of 1.0 means there is no false negatives), F_1-Score and F_2-Score as 90.81 and 86.39, respectively, which represents our model outperforms other existing deep learning models.

Table 3. Segmentation results based on precision and Dice similarity Coefficient

Model	Precision	Dice Score
Ngoc-Quang Nguyen et.al.[9]	-	0.896
Ashkan Tashk et.al.[7]	62.00	-
Le Thi Thu Hong et al. [11]	86.28	0.891
Hemin Ali Qadir et al. [12]	88.35	-
Proposed SU-NET	89.21	0.895

Table 3 demonstrates the performance of various deep learning models on the benchmark CVC-ColonDB Dataset. When the proposed model evaluated using precision and segmentation accuracy as metrics, our model outshined all the existing deep learning models with respect to the Dice similarity coefficient. In summary, the proposed enhanced model has achieved much accurate results compared to all other current models on the whole.

5. Conclusion and Future work

Automated and accurate colorectal polyp segmentation can be rehabilitated by the present diagnosis system for early-stage CRC diagnosis. This paper, proposed a novel SU-NET model for real-time automated colorectal polyp detection and segmentation with good accuracy. The main advantage of the SU-NET model over the other existing deep learning models is that it utilizes an Encoder-Decoder-based neural network that can distinguish and localize the polyps from the CVC-ColonDB database in a fully automated manner. Using an augmented image dataset is in the training phase can reduce the overfitting problem. After training the model, it can detect and segment the polyps accurately in images. It gives a higher performance in terms of Recall (91.3%), F_1-Score (90.81%), F_2-Score (86.39%), Precision (89.21), and Dice similarity coefficient of 0.895 than the other previously proposed deep learning models. However, by considering the quantity of time consumed by the model's training phase, our model still has some flaws. Hence, in the future development of deep learning models, we intend to boost the models' performance by employing accurate polyp detection. Our speculative outcomes showed the superiority of our proposed model over the state-of-the-art polyp segmentation models.

REFERENCES

R. L. Siegel *et al.*, “Colorectal cancer statistics, 2017,” *CA. Cancer J. Clin.*, vol. 67, no. 3, pp. 177–193, May 2017, doi: 10.3322/caac.21395.

“Colon cancer: Symptoms, treatment, and causes.” https://www.medicalnewstoday.com/articles/150496 (accessed Jun. 18, 2020).

S. Rathore, M. Hussain, A. Ali, and A. Khan, “A recent survey on colon cancer detection techniques,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 10, no. 3. pp. 545–563, 2013, doi: 10.1109/TCBB.2013.84.

F. Deeba, F. M. Bui, and K. A. Wahid, “Computer-aided polyp detection based on image enhancement and saliency-based selection,” *Biomed. Signal Process. Control*, vol. 55, p. 101530, Jan. 2020, doi: 10.1016/j.bspc.2019.04.007.

J. Bernal, J. Sánchez, and F. Vilariño, “Towards automatic polyp detection with a polyp appearance model,” in *Pattern Recognition*, Sep. 2012, vol. 45, no. 9, pp. 3166–3182, doi: 10.1016/j.patcog.2012.03.002.

S. Sornapudi, F. Meng, and S. Yi, “Region-based automated localization of colonoscopy and wireless capsule endoscopy polyps,” *Appl. Sci.*, vol. 9, no. 12, p. 2404, Jun. 2019, doi: 10.3390/app9122404.

A. Tashk, J. Herp, E. Nadimi, † Maersk, and M.-K. Muller, “Automatic Segmentation of Colorectal Polyps based on a Novel and Innovative Convolutional Neural Network Approach.” Accessed: Dec. 14, 2020. [Online]. Available: https://ml.sdu.dk/.

J. Kang and J. Gwak, “Ensemble of Instance Segmentation Models for Polyp Segmentation in Colonoscopy Images,” *IEEE Access*, vol. 7, pp. 26440–26447, 2019, doi: 10.1109/ACCESS.2019.2900672.

N. Q. Nguyen and S. W. Lee, “Robust Boundary Segmentation in Medical Images Using a Consecutive Deep Encoder-Decoder Network,” *IEEE Access*, vol. 7, pp. 33795–33808, 2019, doi: 10.1109/ACCESS.2019.2904094.

“IEEE Xplore Full-Text PDF:” https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9163355 (accessed Dec. 14, 2020).

L. T. Thu Hong, N. Chi Thanh, and T. Q. Long, “Polyp Segmentation in Colonoscopy Images Using Ensembles of U-Nets with EfficientNet and Asymmetric Similarity Loss Function,” in *Proceedings - 2020 RIVF International Conference on Computing and Communication Technologies, RIVF 2020*, Oct. 2020, pp. 1–6, doi: 10.1109/RIVF48685.2020.9140793.

H. A. Qadir, Y. Shin, J. Solhusvik, J. Bergsland, L. Aabakken, and I. Balasingham, “Toward real-time polyp detection using fully CNNs for 2D Gaussian shapes prediction,” *Med. Image Anal.*, vol. 68, p. 101897, Feb. 2021, doi: 10.1016/j.media.2020.101897.

K. Simonyan and A. Zisserman, “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION,” 2015. Accessed: Jan. 02, 2021. [Online]. Available: http://www.robots.ox.ac.uk/.

Y. Wu and K. He, “Group Normalization,” *Int. J. Comput. Vis.*, vol. 128, no. 3, pp. 742–755, Mar. 2020, doi: 10.1007/s11263-019-01198-w.

J. Bjorck, C. Gomes, B. Selman, and K. Q. Weinberger, “Understanding Batch Normalization.”

A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier Nonlinearities Improve Neural Network Acoustic Models,” 2013.

O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9351, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017,

doi: 10.1109/TPAMI.2016.2644615.

M. Abadi *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems." Accessed: Nov. 16, 2020. [Online]. Available: www.tensorflow.org.