IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 February 2024, accepted 19 April 2024, date of publication 24 April 2024, date of current version 1 May 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3392947*

RESEARCH ARTICLE

# ContourNet–An Automated Segmentation Framework for Detection of Colonic Polyps

**SAMEENA PATHAN[1], YASHODHARA SOMAYAJI[1], TANWEER ALI[2], (Senior Member, IEEE), AND MODHA VARSHA[1]**

[1]Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India
[2]Department of Electronics and Communication, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India

Corresponding author: Tanweer Ali (tanweer.ali@manipal.edu)

**ABSTRACT** One of the most prevalent forms of malignant tumor affecting the digestive system is the Colorectal Cancer (CRC). Due to the recurrent nature of occurrence of CRC tumors the morbidity and mortality rate associated with CRC tumors is very high, making it the fourth leading cause of cancer worldwide. Although colonoscopy, is considered as pre-dominant screening mechanism for detection of CRC, over the recent years several researchers have attempted in developing AI diagnostic tools for segmenting colonic polyps. However, manual intervention and decreased accuracy rate is a major drawback witnessed by these approaches. In this study, we propose a novel automated colonic polyp segmentation mechanism using U-Net and chroma based deformable model termed as ContourNet. The proposed model eliminates the need for manual identification of region of interest irrespective of the high degree of variability between the lesion and non-lesion pixels. The model considers the chromaticity and statistical information from the identified region of interest to evolve the contour close to the affected area. The proposed algorithm was developed using Clinic CVC dataset for training, and a test accuracy of 94% was obtained. The generalization ability of the proposed design is validated on three different datasets, considering all the images as test set, a high degree of accuracy was obtained in identifying the affected regions. Thereby the results obtained prove the ability of the proposed method to be adopted in developing a CAD tool in detection of pathologies associated with gastroenterology in contrast to the state of art methods reported in literature.

**INDEX TERMS** Colorectal cancer, convolutional neural network, image augmentation, color space augmentation, chroma based deformable model, semantic segmentation.

## I. INTRODUCTION

The large intestine or the colon of a human body is mainly affected by the overgrowth of harmful cells termed as Colorectal Cancer (CRC). Globally, among women it is fourth most prevent cancer and ranks as third most prevalent cancer among men, thus highlighting the demand for early prognosis [1]. Non-cancerous growth of cells on the lining of the colon are considered as primary source of CRC [2]. One of the major ways to prevent this cancer from spreading to the lymph nodes, also considered as an advanced stage of the cancer, would be identification and localization in the earlier stages, thereby reducing the mortality rates and increase in life expectancy among patients. Individuals prone to smoking, drinking alcohol, and elderly people with unhealth lifestyle are susceptible to colonic polyps [3]. Physicians consider these risk factors as major factors in identification of people who are at a risk of developing colonic polyps/CRC. The prevalence of typical symptoms like anemia, rectal bleeding, and abdominal pain highlights the importance of prompt and precise detection because these symptoms can co-exist with other gastrointestinal problems.

Colonoscopy has traditionally been the gold standard for identifying colonic polyps [4]. Nevertheless, it can be difficult

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

to find these polyps during a colonoscopy, and the examining physician may occasionally overlook them. This is where techniques for computer-aided diagnosis are useful. The main goal of these tools is to automatically identify, categorize, and segment colonic polyps from colonoscopy frames and videos. However, most of these techniques are manual, time-consuming, and subjective in nature. The Figure 1 illustrates the region of formation of colonic polyps, and visual perception via a colonoscope. As, it can be observed from Figure 1 that visual perception and degree of intensity variability between the affected and non-affected region is low, making it potentially difficult for a gastroenterologist to identify the poly, thus leading to mis diagnosis or false diagnosis.

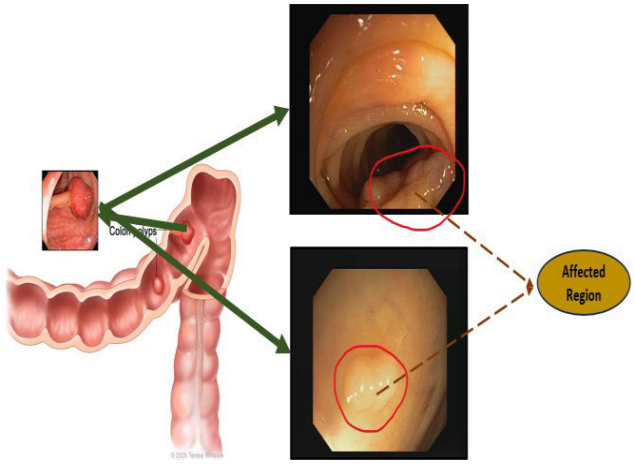


**FIGURE 1.** Illustration of colon polyps in human intestine (Note: Colonoscopic images obtained from CVC-Database [5]).

### A. PATHOLOGICAL SIGNIFICANCE OF SEGMENTATION IN IDENTIFICATION OF COLORECTAL CANCER

Deep learning techniques have had a substantial impact on medical imaging and diagnosis [6], [7]. Convolutional neural networks (CNNs), in particular have become an effective tool for this, providing improved performance in terms of accuracy and speed. The automated analysis of medical images, which includes the detection of colonic polyps, has demonstrated the effectiveness of these neural networks. Moreover, strong network performance has been attained by utilizing transfer learning [8], [9], [10], [11], [12], [13]. By using medical imaging data to fine-tune pre-trained models, this method makes it possible to create extremely accurate diagnostic systems.

Techniques for pre-processing and image augmentation are used to enhance the performance of deep learning models. Through artificially increasing the size and diversity of the training dataset, image augmentation produces models that are more resilient. On the other hand, pre-processing improves image quality and lowers noise, which makes it simpler for the model to precisely identify and segment colonic polyps. The efficiency and efficacy of colorectal cancer screening could be greatly enhanced by these AI-assisted systems, which could ultimately lead to better patient outcomes and possibly even save lives through early detection and treatment. The segmentation of colonic polyps aids in identification of affected region, and aid in disease prognosis, Additionally, localizing the affected region aids in determining the severity level of CRC cancer, when analyzed further. Thus, delineation of region of interest plays a major role in developing a computer assisted diagnostic system.

### B. RELATED WORK

Most of the previous work in polyp classification uses supervised learning processes like a Hessian filter and Support vector machine (SVM), Recurrent Convolutional Neural Network (RCNN), etc. These methods have various demerits such as finding the optimal value of the threshold in the case of SVM, and poor segmentation due to irregularly shaped polyps for SA-DOVA [4]. In [4], the authors adopt an adaptive Markov random field (MRF) method which is an unsupervised learning method. The method over-segments the image into super pixels which are then refined by Local Binary Pattern (LBP) and color features used in the adaptive MRF. Images from CVC-ClinicDB dataset consisting of 612 images each of size $288 \times 384 \times 3$ were used for the model. The method implemented by the authors of [4] can localize the regions which contain polyps and gives a dice value score of 60.77%. In [5], Tashk, Ashkan et al. use a modified CNN network named U-Net for image segmentation. Their model provides an improved semantic segmentation ability for polyp images obtained from colonoscopy. The images for training the model were obtained from openly available datasets (CVC-ColonDB, ETIS-Larib, CVC-ClinicDB). The authors of [5] first extract patches with different rotations from the training dataset. These patches eliminate the need for padding. In [6] the authors have used ''MASK R-CNN'' a neural network used for object detection to detect and segment pictures obtained from a colonoscopy procedure. In [6], the authors combine two MASK R-CNN models (ResNet101 and ResNet50) to enhance performance and improve accuracy. The MASK R-CNN was trained using the intestinal polyps dataset (CVC-ColonDB, ETIS-Larib, CVC-ClinicDB). The method implemented in [6] gives IoU score of 66.06%. In [14], Akbari et al. use an FCN-8s network with otsu thresholding on CVC-ColonDB database and obtain segmentation masks with an accuracy of 97.70%, sensitivity of 74.80% and a dice score of 81%. In [15] Yang et al. used an MRCNN with precise region of interest (PrROI) pooling for the detection and segmentation of colonic polyps. The authors first filter out low quality data using an image filter and then group the images into multiple groups based on the patients and then train the MRCNN model. The model in [15] can detect polyps with an average precision (AP) of 76% and can segment polyps with an IoU score of 86.87%.

The proposed work in this paper contributes the idea of using multiple color spaces and CLAHE augmentation to the images which helps increase the segmentation ability of the model. The proposed model has various merits over previous

works with an IoU score of 89.32% and can even successfully detect small polyps due to the contrast enhancing CLAHE augmentation. It has significantly higher IoU and dice scores and due to the low processing time of 33.44 ms per image can localize polyps and segment them in real-time and hence is a greater asset to clinicians than previous models. Although, significant contributions have been made by the previously reported studies, there exists a gap between the current state and desired state in endoscopic image diagnosis. The pitfalls in this area are given below:

- Although, colonic polyps are a hallmark in detection of colorectal abnormalities, the detection of colonic polyps is difficult to low variability between the different pathologies.
- Lack automated segmentation methods.
- Appropriate choice of initial contours for active contours is challenging.

The objective of the proposed work is to design a novel automatic system for detection of colonic polyps from endoscopic images. The proposed methodology aims to address the aforementioned research gaps associated with the state of art methods.

### C. CONTRIBUTIONS

This paper presents an automated framework for detection of colonic polyps for severity analysis of CRC. Initially, the magnitude of the chroma component of the endoscopic image is extracted, which is further subjected to encoder decoder mechanism to obtain the initial mask. Further, a colonic polyp specific speed function is defined by considering the statistical properties of the polyps to delineate the region of interest from the background. The main contributions of the work are as follows:

- Development of an efficient U -net encoder decoder based on chromaticity features for extraction of initial contour
- Automated evolution of the initial contour to achieve fine-tuned segmentation by minimizing the proposed speed function using stochastic gradient. The proposed Contour Net is automated in nature, and takes into account the color properties of the colonic polyps to define the force function towards the affected region, in spite of the low variabilities between the pixels.
- Generalization ability of the proposed model is tested on unseen data test data (three datasets), achieving an IoU of 94% on Clinic CVC dataset. The proposed algorithm reports highest degree of IoU score in contrast to the state of art approaches reported in literature tested on well-known datasets

## II. PROPOSED SYSTEM

The proposed system comprises of three steps as illustrated in Figure 2. Initially, the endoscopic image is subjected to pre-processing, the second step involves detection of initial contour detection using proposed U net model, followed by chroma based deformable model to generate segmented masks. The aforementioned steps are described below.

### A. DATASET DESCRIPTION

Deep learning models require a vast amount of data for stable training; however, such a large amount of colonoscopy images and their associated ground truth is difficult to obtain, hence there is a need for data augmentation. Images from CVC-ClinicDB were used to train the model [13]. CVC-ClinicDB contains images of 612 polyps sized 288 × 384×3. The images obtained from CVC-ClinicDB were encoded in Tagged image file format (.tif) which was encoded by the image source. The.tif images were found to be incompatible with tensorflow models. Tensorflow models expects the image shape to be in the format of (height, width, channel), however the images were encoded such that the color channel information could not be read by the image handling libraries like OpenCV and matplotlib. Thus, the images had to be converted to.png format to ensure compatibility. Three datasets were used for testing, namely CVC ColonDB, CVC-300, and Kvasir, CVC Colon DB [31] consists of 380 images. Kvasir dataset consists of 1000 images, of varying dimensions, the test images were resized to 350 × 350 to maintain uniform dimensions during testing.

### B. PREPROCESSING

Due to the small number of images, the model would exhibit poor results, and hence data augmentation had to be done to the dataset to increase the number of images. This was done through augmentations such as vertical and horizontal flips, shift, scale, and rotational augmentations. The shift, scale and rotate limits were set to 0.1, 0.5 and 0.2 respectively. The polyp frames were then evaluated in three color spaces, RGB, HSV, and CIELAB [16]. The RGB color space represents each pixel as an array of red, green, and blue hues mixed together with no separate values for saturation and luminance. HSV color space represents images as hue, which represents the shade of the color, saturation, which represents the amount of hue, and value which represents the brightness of colors. In L*a*b* (CIELAB) color space, the (L) lightness channel represents information about the brightness of the image while the 'a' and 'b' channels represent information about the color. The image is then split into its channels and Contrast Limited Adaptive Histogram Equalization (CLAHE) [7], [17] is applied to the 'L' channel to perform image contrast enhancement. CLAHE is an image equalization technique that is used to improve the contrast of the image. Instead of using the entire image, CLAHE operates on small portions of the image called tiles and blends tiles together using bilinear interpolation as illustrated in Figure 3.

### C. INITIAL CONTOUR DETECTION

Various segmentation models were surveyed including the various version of the YOLO algorithm, Fully Convolutional Networks (FCN), ResNet, DenseNet, Mask R-CNN,

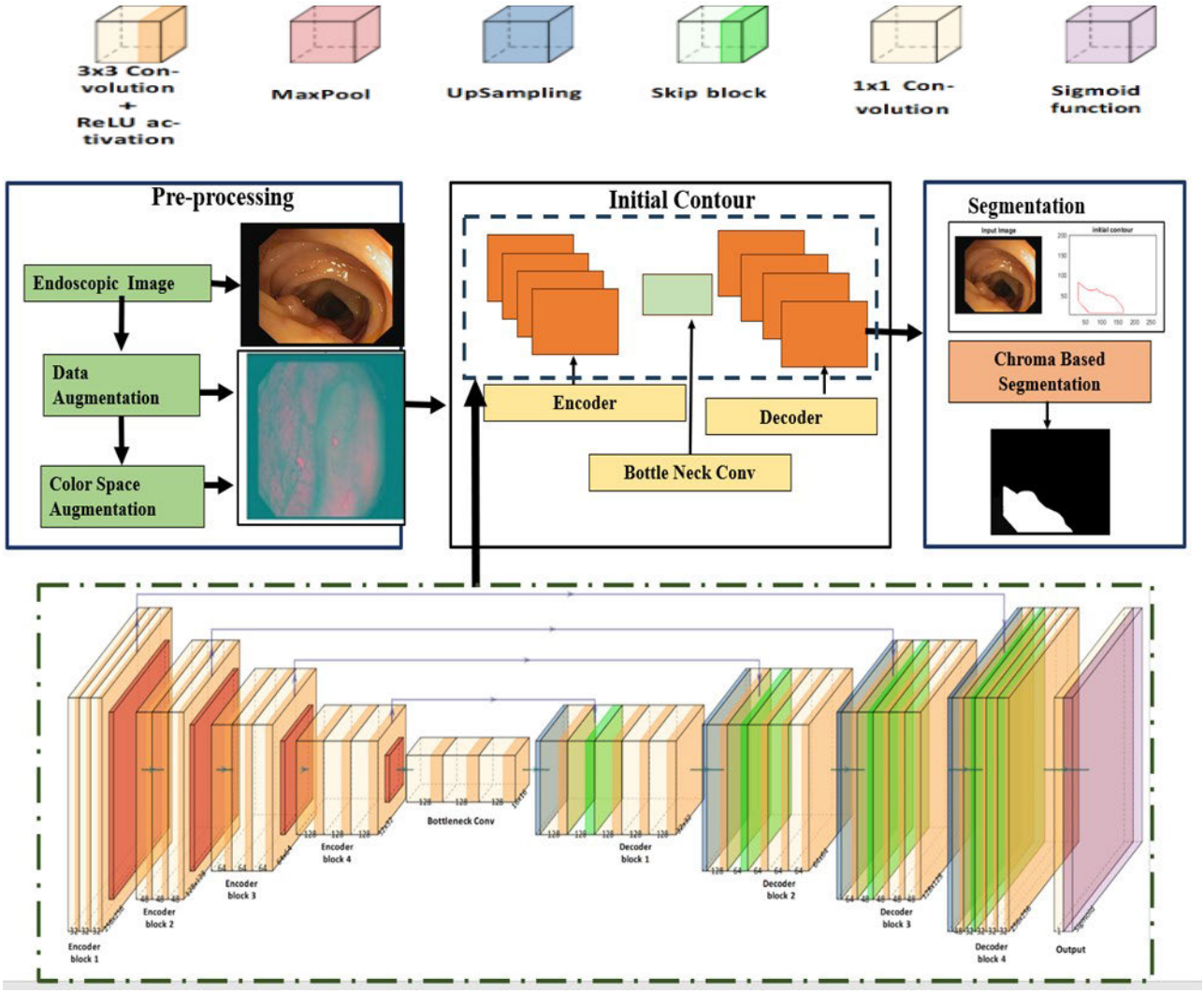


**FIGURE 2.** Overview of the proposed methodology.

U-Net [19] etc. YOLO algorithms have excellent real time performance and was initially developed for automated driving and object detection on the road. Mask R-CNN is a state of the art model used for instance image segmentation. U-Net is an encoder-decoder model that was specifically designed for medial image processing. U-Net has several advantages over traditional FCNs and even Mask RCNN, but its ability to operate more quickly and be compatible with a variety of input dimensions is its most important advantage. One feature that sets U-Net apart from other FCNs has to do with its unique architecture. U-Net is symmetric and can be broken down into three independent pieces known as, down sampling /contracting path, the Bottleneck, and up sampling/expanding path. U-Nets, as opposed to FCNs, have skip links in between pathways that downscale and upscale data. This structure protects U-Nets against overfitting because of the insufficient number of data store pictures. As the dataset contained very few images, image augmentation was essential to obtain satisfactory results. Some of the augmentation applied include, vertical and horizontal flips, image padding, shift, scale, and rotate augmentation which shifts the image, changes the zoom level and rotates the image. For the first trial run, the user implemented U-Net model was trained on the images directly without augmentation.

The model contained four encoder/decoder blocks. Each encoder/decoder block contained two convolution blocks with batch normalization and ReLU activation function applied to each block. The proposed U-Net model contains four blocks each in the down sampling and up sampling path with filters increasing from 32 for the initial layer, 48 for the second layer, 64 for the penultimate layer, and 128 for the final layer. Each block contains three sets of $3 \times 3$ convolution, batch normalization, and ReLU activation function with a max pooling operation after each block and skip links between blocks in the contacting and expanding path. The model uses sigmoid function to output the generated mask at the end of an iteration. The input images are first resized to $256 \times 256$ pixels. Traversal through each layer after max

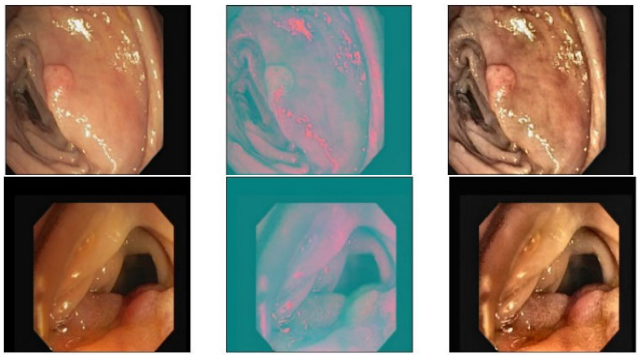


**FIGURE 3.** Illustration of pre-processing (a) represents the original polyp image, (b) represents the converted LAB image, and (c) represents the image after CLAHE is applied.

**TABLE 1.** Model architecture description.

| Method | Layers | Description |
|---|---|---|
| Down sampling | First Layer (32 Filters) | 3 sets 3x3 Convolution, Batch Normalization, ReLU |
| | Second Layer (48 Filters) | |
| | Third Layer (64 Filters) | |
| | Fourth Layer (128 Filters) | |
| U sampling | First Layer (32 Filters) | 3 sets 3x3 Convolution, Batch Normalization, ReLU |
| | Second Layer (48 Filters) | |
| | Third Layer (64 Filters) | |
| | Fourth Layer (128 Filters) | |
| Expanding Layer | | Sigmoid Activation Function |

pooling reduces the image to half its input size, thus the image is down sampled. In the expanding layer, the image is rebuilt using information from skip links and the input down sampled image, and the binary segmented mask is generated. The model architecture is depicted in Table 1.

### D. METHOD FOR CHOOSING THE PARAMETERS

The model was run for 50 epochs and Adam optimizer was used to dynamically change the learning rate. For a batch Size of 8, Learning rate of 10-4. After 50 epochs the training IoU and loss was 0.6678 and 0.2011 while the validation IoU and loss was 0.5171 and 0.3178. Each Epoch took 9s when trained on google colab. When the model was evaluated on unseen test data, the IoU was 0.5172 and Loss was 0.3182. The output masks were often incomplete and some polyps were missed altogether. For the second trial run, the convolutional blocks in each encoder/decoder block was increased from 2 to 3 anticipating better feature extraction by increasing the complexity of the model. However, the model was found to be worse than the previous trial. The model was run for 50 epochs and Adam optimizer was used to dynamically change the learning rate. For a batch Size of 8, Learning rate of $10^{-4}$. After 50 epochs the training IoU and loss was 0.5297 and 0.3106 while the validation IoU and loss was 0.3918 and 0.4354. Each Epoch took 13s when trained on google colab. When the model was evaluated on unseen test data, the IoU was 0.4159 and Loss was 0.4130. The model was worse when the convolutional layers were increased. The Table 2 provides the details regarding the parameters chosen.

**TABLE 2.** Values for network parameters.

| Network Parameters | Values |
|---|---|
| Image size | 256x256 |
| Batch size | 8 |
| Epochs | 60 |
| Learning rate | 0.0001 |
| Weight Decay | 0.01 |
| Optimizer | Adam |

For the third trial run, data augmentation was done which increased the number of images from 612 to 1224. The model was run for 50 epochs and Adam optimizer was used to dynamically change the learning rate. IoU and loss was 0.7836 and 0.1161. Each epoch took 25s when trained on google colab. When the model was evaluated on unseen test data, the IoU was 0.7949 and Loss was 0.1177. Increasing the number of images vastly improved the output of the model. For the fourth trial run, data augmentation was done which increased the number of images from 612 to 2448. After 50 epochs the training IoU and loss was 0.9346 and 0.0339 while the validation IoU and loss was 0.8682 and 0.0689. Each epoch took 49s when trained on google colab. When the model was evaluated on unseen test data, the IoU was 0.8759 and Loss was 0.0664. Doubling the images from the previous run provided significant improvement in results while doubling the training time. For the fifth trial run the number of epoch was reduced to 30 as there were marginal improvements in the subsequent epochs. The model was run for 30 epochs and Adam optimizer was used to dynamically change the learning rate. Subsequently, there were eleven trials, for obtaining the optimized values. The Figure 4 provides an illustration of the IoU results obtained. For the eleventh trial, CLAHE was applied to the L channel for LAB images. The performance values for each of the iterations with respect to the ground truths is given in Table 3. Since, CIELAB with CLAHE provided better performance in contrast to the other planes, the CIELAB+CLAHE was used further for generation of initial contours for the training set. The Algorithm 1 steps provides a brief overview of the steps carried out to obtain initial contour close to the region of interest.

**TABLE 3. Comparison of results for different color spaces.**

| Trial | IoU score | Dice Loss |
|---|---|---|
| Model with RGB color space | 0.85 | 0.078 |
| Model without CLAHE augmentation in L*a*b* color space | 0.86 | 0.075 |
| Model with HSV color space | 0.87 | 0.063 |
| Model with CLAHE augmentation in L*a*b* | 0.89 | 0.057 |

### E. CHROMA BASED ACTIVE CONTOUR FOR SEGMENTATION

The accuracy of detection of the region of interest (ROI) mainly relies on the initial contour selected and the speed function evolution. Hence, rather than choosing some random points to determine the initial contour, the U-net model is used to determine the initial contour such that initial contour is placed in close proximity to the region of interest. The initial contour is defined as given in (1).

$$\emptyset\,(x, y, 0) = \left\{ \begin{array}{l} 1 \; if \; F\,(x, y) \cong I\,(x, y) \\ 0 \; if \; F\,(x, y) \mid \cong I(x, y) \end{array} \right\} \tag{1}$$

where, $F(x, y)$ is the input image and $I(x, y)$ is the data used for training the proposed U net model. Since lab color space considers the perceptual differences between the two colors, the magnitude of the chroma component is used to generate the input image. It can be observed from the Figure 3 (b), that the chroma image provides better visual perception of ROI in contrast to the RGB image. The image is obtained using (2).

$$F'\,(x, y) = sqrt\left(F\,(x, y)\,(a)^2 + F\,(x, y)\,(b)^2\right) \tag{2}$$

Here, $F(x, y)(a)$ and $F(x, y)(b)$ are the two components of the CIELAB color space. The speed function is given by (3)-(5).

$$F = \left(|(F_R\,(x, y) - u_R|^2 - |F_R\,(x, y) - v_R|^2\right) + \left(|F_{NR}\,(x, y) - u_{NR}|^2 - |F_{NR}\,(x, y) - v_{NR}|^2\right) \tag{3}$$

where,

$$F_R\,(x, y) = e \wedge -\frac{\left(F'^{(x,y)} - \mu_R\right)^2}{2\sigma_R^2} \tag{4}$$

$$F_{NR}\,(x, y) = e \wedge -\frac{\left(F'^{(x,y)} - \mu_{NR}\right)^2}{2\sigma_R^2} \tag{5}$$

where, $\mu_R$ and $v_R$ are the mean values of $F_R\,(x, y)$ within and outside the initial contour with respect to the region of interest. Similarly, $\mu_{NR}$ and $v_{NR}$ are the mean values of $F_{NR}\,(x, y)$ within and outside the initial contour with respect to the non-region of interest. The speed function is minimized to reach the global minimum by using stochastic gradient decent. The initial contour is overlayed on the endoscopic image to determine the statistical values of the parameters with respect to both the regions. The Algorithm 2, provides the detailed steps for segmentation using chroma based deformable model.

**Algorithm 1**

Step 1: Collection of data and data augmentation
Step 2: Conversion to CIELAB color space
Involves conversion of image from RGB to XYZ(1) and XYZ to LAB(2) [16]
Inverse companding of V ∈ {R,G,B} to v ∈ {r,g,b}
Inverse gamma companding: v=$V^{\gamma}$ where $\gamma = 2.2$ for sRGB
Inverse sRGB companding: $v = \left\{ \begin{array}{l} \frac{V}{12.92} \; if \; V \leq 0.04045 \\ \left(\frac{V+0.055}{1.055}\right)^{2.4} \; otherwise \end{array} \right\}$
Inverse L∗ companding: $v = \left\{ \begin{array}{l} \frac{100v}{k} \; if \; V \leq 0.08 \\ \left(\frac{V+0.16}{1.16}\right)^{3} \; otherwise \end{array} \right\}$

$$k = \left\{ \begin{array}{l} 903.3 \; Actual \; CIE \; standard \\ \frac{24389}{27} \; Intent \; of \; CIE \; standard \end{array} \right\}$$

Linear RGB to XYZ:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [M] \begin{bmatrix} r \\ g \\ b \end{bmatrix}$$

Where M is 3 × 3 matrix, $M = \begin{bmatrix} 0.4124 & 0.3576 & 0.1804 \\ 0.2127 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.95036 \end{bmatrix}$ for sRGB
Step 3: Conversion from XYZ to LAB requires a reference white $(X_r, Y_r, Z_r)$ $L = 116 f_y - 16$, $a = 500\,(f_x - f_y)$, $b = 200\,(f_y - f_z)$
where, $f_x = \left\{ \begin{array}{l} \sqrt[3]{x_r} \; if \; x_r > \epsilon \\ \frac{kx_r+16}{116} otherwise \end{array} \right\}$,
$f_y = \left\{ \begin{array}{l} \sqrt[3]{y_r} \; if \; y_r > \epsilon \\ \frac{ky_r+16}{116} otherwise \end{array} \right\} f_z = \left\{ \begin{array}{l} \sqrt[3]{z} \; if \; z_r > \epsilon \\ \frac{kz_r+16}{116} otherwise \end{array} \right\}$, where,
$x_r = \frac{X}{X_r}, y_r = \frac{Y}{Y_r}, z_r = \frac{Z}{Z_r}$
$k = \left\{ \begin{array}{l} 903.3 \; Actual \; CIE \; standard \\ \frac{24389}{27} \; Intent \; of \; CIE \; standard \end{array} \right\}$
$\epsilon = \left\{ \begin{array}{l} 0.008856 \; Actual \; CIE \; standard \\ \frac{216}{24389} \; Intent \; of \; CIE \; standard \end{array} \right\}$
Step 4: Apply CLAHE to the 'L' channel of the image
If x is a sample image, then,
(L, a, b) = split (x)
L′ = CLAHE(L)
x′ = merge(L′, a, b)
Step 5: Train the model with images from the training set as input
Step 6: Obtain the initial contour from the last layer

## III. RESULTS

### A. METRICS FOR EVALUATION

The model is evaluated on the Intersection over Union (IoU) score and Dice loss. IoU also known as Jaccard Index is a measure of similarity between sets. In semantic segmentation and object detection tasks, IoU is used to measure the overlap that exists between the ground truth masks and the generated mask. IoU is defined as (6).

$$IoU = \frac{(A \cap B)}{(A \cup B)} \tag{6}$$

**Algorithm 2** Segmentation Using Chroma Based Deformable Model

Input: Endoscopic Image$F(x,y) = \{R(x,y), G(x,y), B(x,y)\}$ ~CIELAB( $F(x,y)$

Output: Segmented Image $S(x,y)$

1. Initialization
   At$t = 0$; $t = Iteration$
   Compute $\emptyset(x,y,0)$ usingequation (1) ;
2. Obtain the chroma component using CIELAB( $F(x,y) = F'(x,y)$ using (2)
3. Curve Evolution
   At $t = t+1$
   Compute $F$ using equation (2);
   Compute the gradient with respect to $F_R(x,y)$ and non region of interest $F_{NR}(x,y)$ using (4)-(5)
   where, *Mean of the polyp area* : $\mu_R = \frac{\sum I_R(x,y)}{N}$
   where, $I_R(x,y)$ = ROI Area obtained by overlaying Initial Contour on the endoscopic Image
   N= Total Number of polyp pixels
   *Standard Deviation of the Polyp area* : $\sigma_R = \sqrt{\frac{1}{N}\sum(I_R(x,y) - \mu_R)}$
   *Mean of the non polyp area* : $\mu_{NR} = \frac{\sum I_{NR}(x,y)}{N}$
   where, $I_{NR}(x,y)$ = NROI Area obtained by overlaying Initial Contour on the endoscopic Image
   N= Total Number of nonpolyp pixels
   *Standard Deviation of the Polyp area* : $\sigma_{NR} = \sqrt{\frac{1}{N}\sum(I_{NR}(x,y) - \mu_{NR})}$
   *Minimize F using Stochastic Gradient Descent*
4. Stopping criteria: min (*SGD*)
   Constant Contour Position
   or
   $t > t_{max}$, where $t_{max} = 200$

where A and B are the set of predicted values, and the ground truth values respectively. The value of IoU varies from 0 to 1 where 0 indicates that there is no overlap in the two images and 1 indicates that the two images overlap perfectly. The dice coefficient is an index that measures the pixel level overlap between the generated masks and the ground truth masks. Dice coefficient and IoU are positively correlated. Dice loss is the preferred loss function while evaluating the results of a semantic segmentation model. It is defined as given in (7).

$$Dice\ Coefficient = \frac{2 * |A \cap B|}{|A| + |B|} \tag{7}$$

where A is the set of predicted values and B is the ground truth values. The dice loss is defined as in (8).

$$Dice\ loss = 1 - Dice\ Coefficient \tag{8}$$

The value of the dice coefficient varies from 0, which indicates no spatial overlap between the two sets, to 1 which indicates complete overlap between the two sets. The model should aim to minimize the dice loss and maximize the IoU. As IoU and Dice loss is not available as a pre-built metric in TensorFlow, a custom IoU and Dice loss metric was implemented.

### B. RESULTS USING U NET SEGMENTATION

The proposed model was trained for 60 epochs with batch size set to 8 and initial learning rate set to 0.0001. An increase in the batch size to 16 led to poorer results and 8 was found to be optimal. Adam optimizer [20] was used to optimize the learning rate and reduce loss with weight decay set to 0.01, further exponential moving averages (EMA) were used. EMA computes the exponential moving average of the weights and periodically replaces the weights with its moving average. At the end of training the model had a validation IoU score of 0.88 with dice loss of 0.06. The model was tested on 224 images which were not used for training, the proposed model exhibited an IoU score of 0.8932 (89.32%) and a dice loss of 0.0578 (5.78%). The Figure 4 illustrates the results obtained using the efficient model.

### C. RESULTS USING U NET SEGMENTATION WITH CHROMA BASED DEFORMABLE MODEL

The model was tested on 224 images which were not used for training, the proposed model exhibited an IoU score of 0.94 (94.22%) and a dice loss of 0.0366 (3.66%). The Figure 5 illustrates the segmentation results obtained for the initial contour. The Figure 6 shows the results of segmentation obtained ContourNet, wherein it can be observed that the detected polyp region and the ground truth overlaps to a greater extent as indicated by zoomed image in Figure 6(e). The Table 4, provides the quantitative results of segmentation. An improved accuracy of 0.94 was obtained by the proposed automated design comprising of U Net and Chroma based deformable model. Few suggested approaches reported in literature have combined may involve modifying the backbone U-Net model to a U-Net++ [21] or ResUNet model [22] etc.

### D. ABLATION STUDY

In this section, the results obtained after conducting ablation studies are highlighted with the goal of analyzing the effectiveness of the proposed modules. In contrast to the U-Net, the proposed ContourNet achieves an improvement in the dice score and IoU score by approximately by 5% on CVC ClinicDB dataset. Similarly, an improvement of 7% in IoU score and 4% in Dice Score was observed for Kvasir dataset using the proposed ContourNet in contrast to standalone U Net model. A complete analysis pertaining to the performance with respect to each of the models in terms of three iterations is illustrated by the boxplots in Figure 7.

The first iteration descries the ablation study carried out by applying the U Net model without preprocessing considering the RGB color space. Considering the CLAHE augmentation in LAB color space, the role of LAB color space is highlighted by the fact that perceptual differences between the colonic polyps and non polyp regions are visibly highlighted. This aids in significantly achieving the automatic positioning of the initial contour towards the region

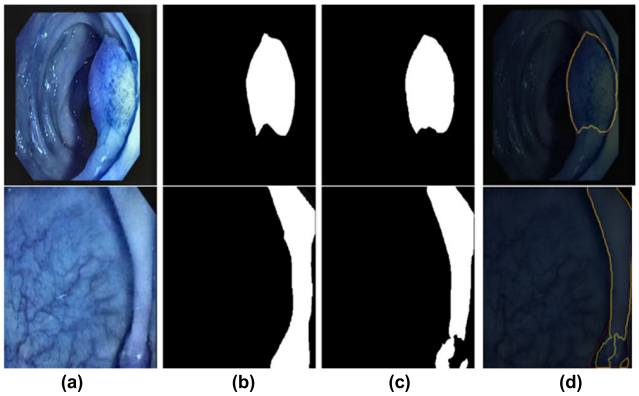


**FIGURE 4.** Column (a) is the input image, column (b) is the ground truth mask, column (c) is the generated mask, and column(d) shows the edge overlap of the ground truth mask and the generated mask over the original image. The red border represents the ground truth mask edge and the yellow border represents the generated mask edge.

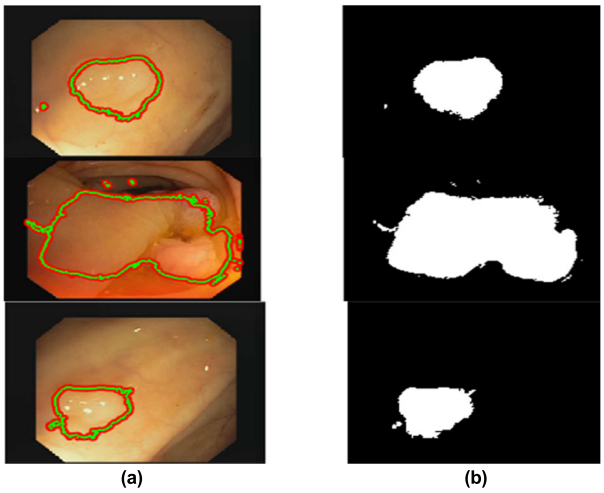


**FIGURE 5.** Results for segmentation, (a) Image with the contour obtained, (b) is the generated mask.

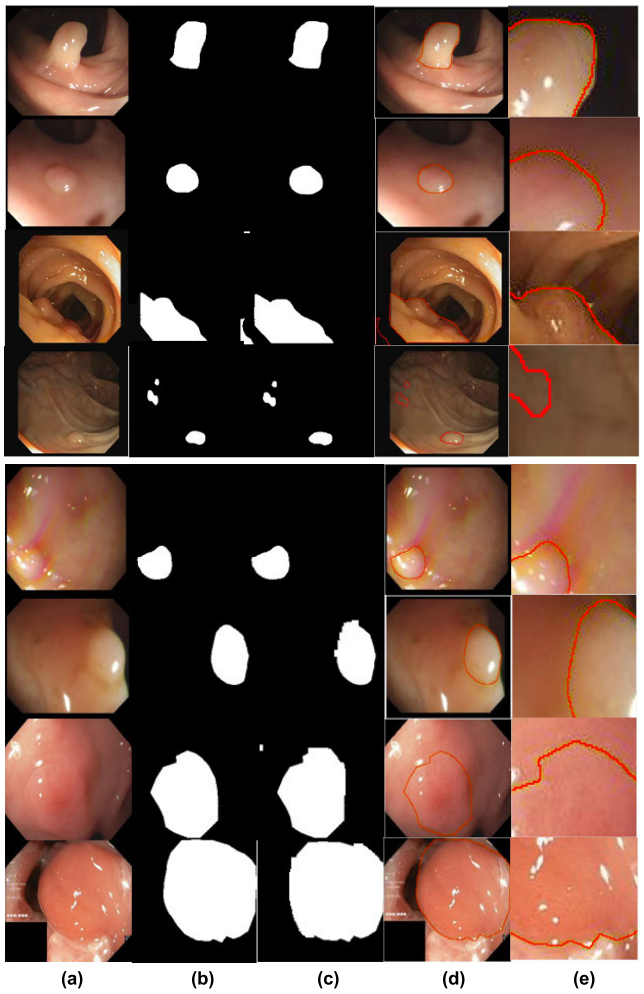


**FIGURE 6.** Results for segmentation (a) Original Image, (b) Ground truth, (c) Segmented Mask, (d) segmented mask and ground truth overlapped on original image, red indicates segmented mask, yellow indicates ground truth, (e) zoomed out image of image given in (d).

of interest. The second iteration describes the use of preprocessing and U net for identification of initial contour. The U net model for initial contour as a standalone model for segmentation of polyp region produces an IoU score of 85% on the well-known Clinic DB dataset. Further, it can also be seen for all the four datasets as illustrated in Figure 7. Additionally, one of the major highlight is the fact that UNet in conjunction with chroma based deformable model (ContourNet) takes into account the border irregularity of the polyp, which is a major criteria studies by gastroenterologists for determining the severity of the colorectal cancer. The gradient of the force function is modified to incorporate the pixel information from ROI and NROI areas, this tends the movement of the contour towards the polyp region, despite the existence of low color variability between the pixels.

Another, major feature of the ContourNet is the use of standard deviation function in computation of the gradient, this further enhances the ability to retain the border irregularity measure of the polyp. The ablation study was carried for the four datasets and the quantitative results obtained for the three iterations are termed as follows:

I: Effect of UNet model without pre-processing.

II: Effect of UNet model with pre-processing.

III: Effect of UNet model with pre-processing and chroma based model (ContourNet).

The results obtained by eliminating one module at a given instant is reported in Table 4 for better insights.

### *E. GENERALIZIBITY AND COMPARATIVE ANALYSIS*

The generalization ability of the proposed ContourNet is validated using three additional datasets namely the Colon DB, CVC-300, and Kvasir dataset. The algorithm was developed using the Clinic CB dataset, and the generalization ability is tested without altering the parametric values for the unseen test data. The performance of the model is given in Table 4 termed as iteration III (highlighted in bold). Additionally, we compared the proposed model with existing state-of-art methodologies developed using the

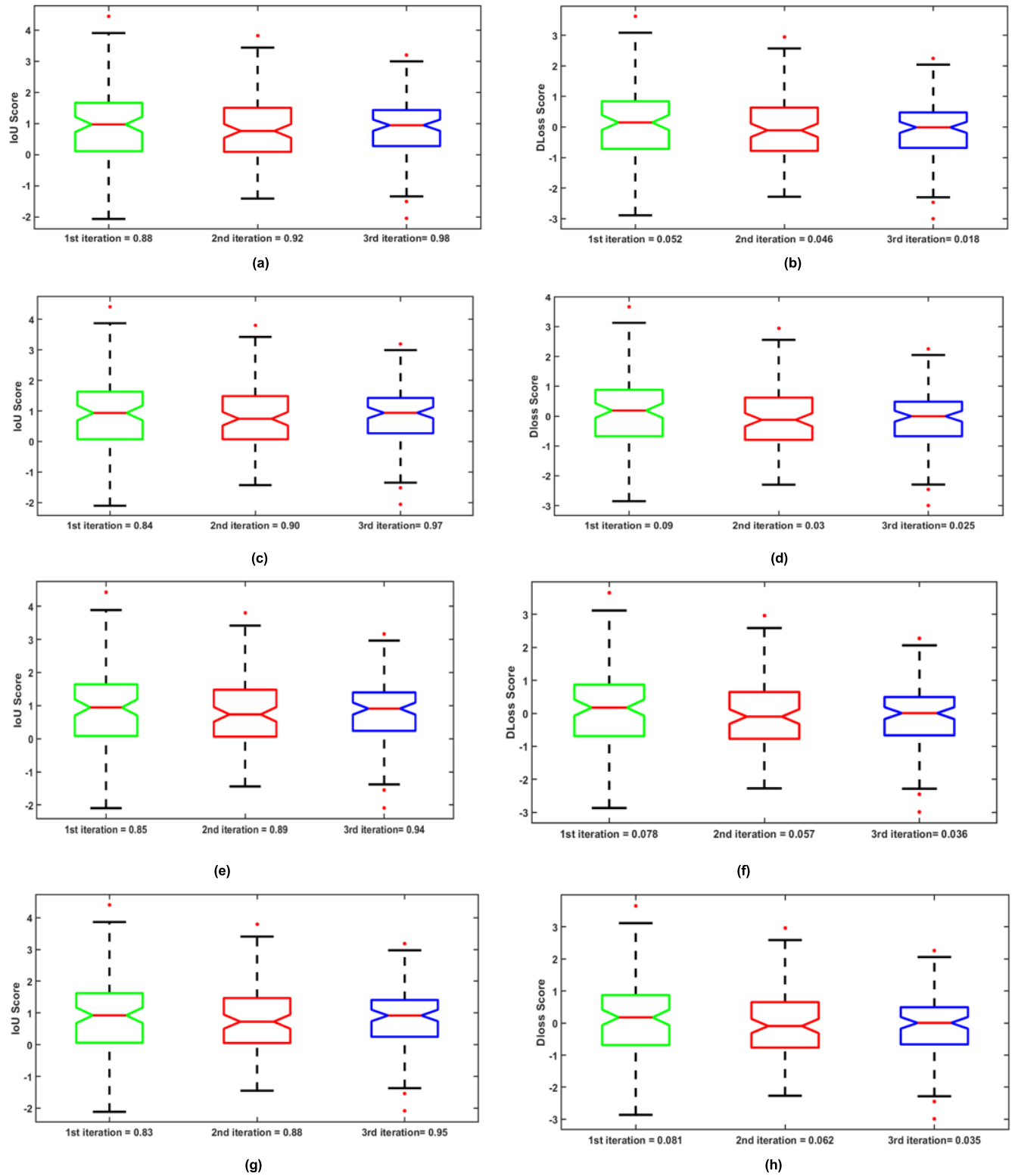


**FIGURE 7.** Box plots for ablation studies for the four datasets, (a,c,e,g) IoU score, (b,d,f,h) Dice Loss Score Row 1: Colon DB, Row 2: CVC 300, Row 3: Clinic DB, Row 4: Kvasir dataset.

respective datasets. As it can be observed form the Table 5. the results obtained using proposed ContourNet model are superior in contrast to the state of art studies reported in literature.

**TABLE 4.** Quantitative results for ablation study.

| Dataset | Colon DB | | | CVC Clinic DB | | | CVC 300 | | | Kvasir | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Iteration | I | II | III | I | II | III | I | II | III | I | II | III |
| IoU | 0.88 | 0.92 | **0.98** | 0.85 | 0.89 | **0.94** | 0.84 | 0.90 | **0.97** | 0.83 | 0.88 | **0.95** |
| Dice Loss | 0.052 | 0.046 | **0.018** | 0.078 | 0.057 | **0.036** | 0.09 | 0.031 | **0.025** | 0.081 | 0.062 | **0.031** |

**TABLE 5.** Quantitative results for the four datasets.

| Dataset | CVC Clinic DB | | CVC Colon DB | | CVC 300 | | Kvasir | |
|---|---|---|---|---|---|---|---|---|
| Performance Metric | IoU | Dice Loss | IoU | Dice Loss | IoU | Dice Loss | IoU | Dice Loss |
| **Reference** | 0.75 | 0.82 | | | | | 0.74 | 0.81 |
| U Net [18] | 0.72 | 0.79 | | | | | 0.74 | 0.81 |
| U Net ++ [21] | 0.84 | 0.89 | | | | | 0.84 | 0.89 |
| PraNet [23] | 0.88 | 0.93 | | | | | 0.82 | 0.89 |
| SA-Net [24] | 0.87 | 0.93 | | | 0.65 | 0.73 | 0.81 | 0.88 |
| CCBANet [25] | 0.87 | 0.92 | | | 0.60 | 0.67 | 0.81 | 0.87 |
| VPS [26] | 0.90 | 0.94 | 0.59 | 0.68 | 0.76 | 0.84 | 0.83 | 0.89 |
| ColonFormer [27] | 0.90 | 0.94 | 0.70 | 0.78 | 0.82 | 0.89 | 0.85 | 0.90 |
| BA-Net [28] | 0.75 | 0.82 | 0.69 | 0.76 | 0.83 | 0.90 | 0.74 | 0.81 |
| PNS-Net [29] | | | 0.57 | 0.64 | | | | |
| SCR-Net [30] | | | 0.47 | 0.56 | | | | |
| **Contour Net** | **0.94** | **0.97** | **0.98** | **0.99** | **0.97** | **0.98** | **0.95** | **0.97** |

## IV. CONCLUSION

This paper describes an automated segmentation framework for detection of colonic polyps in endoscopic images. It starts with preprocessing to obtain color space that is optimal for segmentation of colonic polyps. The color space is obtained by taking the magnitude of the a and b components of CIELAB color space. An optimized encoder and decoder is developed to obtain the initial masks for segmentation using the encoder decoder optimized U net model. The initial masks are further used by the chroma based deformable models to obtain fine tune segmentation. The chroma based deformable model are formulated considering the Gaussian nature of distributions of pixels, hence the speed function is formulated such that information specific to colonic and non-colonic areas in incorporated and hence minimized using stochastic gradient descent.

The proposed system for polyp segmentation eliminates the need for manual border delineation for segmentation of colonic polyps. The proposed model is completely automated in nature, the trained U net model was used to generate masks for test set images (224 images), the test masks were further subjected to fine tune segmentation thereby achieving a segmentation accuracy of 94% on Clinic CVC dataset. Further, similar improved performance was obtained on other three datasets. The model exhibits great performance and can localize polyps and display the generated masks in real-time and hence can be a great asset to clinicians and can help automate the detection of pre-malignant polyps thereby helping in the early detection of colorectal cancer helping save lives. The proposed work can be utilized to further improve classification models by providing the generated segmented images instead of the entire image, improving the time in which the classifier can generate results and improving the accuracy of the classifier. The model can be improved further to obtain better results. Further pre-processing like selection of patches and post-processing may also help improve the results. Thus, future directions include evaluation of the proposed methodologies in developing a computer assisted endoscopic framework for detection of gastrointestinal abnormalities.

## REFERENCES

[1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA: Cancer J. Clinicians*, vol. 65, no. 2, pp. 87–108, Mar. 2015.

[2] M. Taghiakbari, Y. Mori, and D. von Renteln, "Artificial intelligence-assisted colonoscopy: A review of current state of practice and research," *World J. Gastroenterol.*, vol. 27, no. 47, pp. 8103–8122, Dec. 2021, doi: 10.3748/wjg.v27.i47.8103.

[3] J. Pan, L. Cen, L. Xu, M. Miao, Y. Li, C. Yu, and Z. Shen, "Prevalence and risk factors for colorectal polyps in a Chinese population: A retrospective study," *Sci. Rep.*, vol. 10, no. 1, p. 6974, Apr. 2020, doi: 10.1038/s41598-020-63827-6.

[4] P. Sasmal, M. K. Bhuyan, S. Dutta, and Y. Iwahori, "An unsupervised approach of colonic polyp segmentation using adaptive Markov random fields," *Pattern Recognit. Lett.*, vol. 154, pp. 7–15, Feb. 2022, doi: 10.1016/j.patrec.2021.12.014.

[5] A. Tashk, J. Herp, and E. S. Nadimi, "Automatic segmentation of colorectal polyps based on a novel and innovative convolutional neural network approach," *WSEAS Trans. Syst. Control*, vol. 14, pp. 384–391, Dec. 2020.

[6] 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.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Munich, Germany. Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.

[8] Y. Shin, H. A. Qadir, L. Aabakken, J. Bergsland, and I. Balasingham, "Automatic colon polyp detection using region based deep CNN and post learning approaches," *IEEE Access*, vol. 6, pp. 40950–40962, 2019, doi: 10.1109/ACCESS.2018.2856402.

[9] H. Chen, D. Ni, J. Qin, S. Li, X. Yang, T. Wang, and P. A. Heng, "Standard plane localization in fetal ultrasound via domain transferred deep neural networks," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 5, pp. 1627–1636, Sep. 2015.

[10] H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers, "Interleaved text/image deep mining on a large-scale radiology database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–10.

[11] Y. Yao, S. Gou, R. Tian, X. Zhang, and S. He, "Automated classification and segmentation in colorectal images based on self-paced transfer network," *BioMed Res. Int.*, vol. 2021, Jan. 2021, Art. no. 6683931, doi: 10.1155/2021/6683931.

[12] E.-H. Dulf, M. Bledea, T. Mocan, and L. Mocan, "Automatic detection of colorectal polyps using transfer learning," *Sensors*, vol. 21, no. 17, p. 5704, Aug. 2021, doi: 10.3390/s21175704.

[13] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. Saliency maps from physicians," *Computerized Med. Imag. Graph.*, vol. 43, pp. 99–111, Jul. 2015.

[14] M. Akbari, M. Mohrekesh, E. Nasr-Esfahani, S. M. R. Soroushmehr, N. Karimi, S. Samavi, and K. Najarian, "Polyp segmentation in colonoscopy images using fully convolutional network," presented at the Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2018, pp. 69–72.

[15] X. Yang, Q. Wei, C. Zhang, K. Zhou, L. Kong, and W. Jiang, "Colon polyp detection and segmentation based on improved MRCNN," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 4501710, doi: 10.1109/TIM.2020.3038011.

[16] B. Lindbloom. (Jul. 15, 2013). *Useful Color Equations*. Accessed: May 12, 2023. [Online]. Available: http://www.brucelindbloom.com/index.html?Math.html

[17] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Müller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," in *Proc. 1st Conf. Vis. Biomed. Comput.*, Atlanta, GA, USA, 1990, pp. 337–345, doi: 10.1109/VBC.1990.109340.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention* (Lecture Notes in Computer Science), vol. 9351, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[20] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, Apr. 2019, doi: 10.1016/j.isprsjprs.2020.01.013.

[21] Z. Zhou, M. Mahfuzur Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," 2018, *arXiv:1807.10165*.

[22] W. Wang, J. Tian, C. Zhang, Y. Luo, X. Wang, and J. Li, "An improved deep learning approach and its applications on colonic polyp images detection," *BMC Med. Imag.*, vol. 20, p. 83, Jul. 2020, doi: 10.1186/s12880-020-00482-3.

[23] D. P. Fan, G. P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "PraNet: Parallel reverse attention network for polyp segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 263–273.

[24] J. Wei, Y. Hu, R. Zhang, Z. Li, S. K. Zhou, and S. Cui, "Shallow attention network for polyp segmentation," in *Proc. 24th Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Strasbourg, France: Springer, 2021, pp. 699–708.

[25] T. C. Nguyen, T. P. Nguyen, G. H. Diep, A. H. Tran-Dinh, T. V. Nguyen, and M. T. Tran, "CCBANet: Cascading context and balancing attention for polyp segmentation," in *Proc. 24th Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Strasbourg, France: Springer, 2021, pp. 633–643.

[26] G.-P. Ji, G. Xiao, Y.-C. Chou, D.-P. Fan, K. Zhao, G. Chen, and L. Van Gool, "Video polyp segmentation: A deep learning perspective," *Mach. Intell. Res.*, vol. 19, no. 6, pp. 531–549, Dec. 2022.

[27] N. T. Duc, N. T. Oanh, N. T. Thuy, T. M. Triet, and V. S. Dinh, "ColonFormer: An efficient transformer based method for colon polyp segmentation," *IEEE Access*, vol. 10, pp. 80575–80586, 2022.

[28] H. Xia, Y. Qin, Y. Tan, and S. Song, "BA-Net: Brightness prior guided attention network for colonic polyp segmentation," *Biocybern. Biomed. Eng.*, vol. 43, no. 3, pp. 603–615, Jul. 2023.

[29] G. P. Ji, Y. C. Chou, D. P. Fan, G. Chen, H. Fu, D. Jha, and L. Shao, "Progressively normalized self-attention network for video polyp segmentation," in *Proc. 24th Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)* Strasbourg, France: Springer, 2021, pp. 142–152.

[30] H. Wu, J. Zhong, W. Wang, Z. Wen, and J. Qin, "Precise yet efficient semantic calibration and refinement in ConvNets for real-time polyp segmentation from colonoscopy videos," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, 2021, pp. 2916–2924.

[31] *PolypDataSet*. Accessed: Feb. 12, 2024. [Online]. Available: https://figshare.com/articles/figure/Polyp_DataSet_zip/21221579

**SAMEENA PATHAN** is currently an Assistant Professor with the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. Her research interests include pattern recognition, medical image analysis, artificial intelligence, and machine learning.

**YASHODHARA SOMAYAJI** received the B.E. degree in information and communication technology from Manipal Institute of Technology, Manipal, in 2022. His research interests include pattern recognition, medical image analysis, artificial intelligence, and machine learning.

**TANWEER ALI** (Senior Member, IEEE) is currently an Associate Professor with the Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. He is an active Researcher in the field of microstrip antennas, wireless communications, medical imaging, and microwave imaging. He has published more than 130 papers in reputed web of science (SCI) and Scopus indexed journals and conferences. He has led seven Indian patents, of which three have been published. He is on the Board of a Reviewer of journals, such as IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS, IEEE ACCESS, and many other journals. He has been listed in top 2% scientists across the world for the year 2021 and 2022 by the prestigious list published by Stanford University, USA, indexed by Scopus.

**MODHA VARSHA** is currently pursuing the M.Tech. degree with Manipal Institute of Technology, Manipal. Her research interests include pattern recognition, medical image analysis, artificial intelligence, and machine learning.