Journal of Theoretical and Applied Information Technology

15th November 2023. Vol.101. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# FULLY CONVENTIONAL CASCADED MULTI DIMENSIONAL NETWORK FOR COLORECTAL POLYP SEGMENTATION OF COLONOSCOPY IMAGES

**M.N. PRASHANTH[1], S. SENTHILKUMAR [2]**

[1] Research scholar, Department of CSE, Presidency University, Bangalore
[2] Professor, Department of CSE, Presidency University, Bangalore
Email: prashanthphdcse@gmail.com[1] , harisen1234@yahoo.co.in[2]

**ABSTRACT**

Early polyps detection can reduce the risk of colorectal cancer. Segmentation of colonoscopy images can facilitate a faster diagnosis of polyps. However, accurate polyp image segmentation is a difficult process because the size, shape, and location of polyps vary, and expert experience directly affects this process. This research article presents a novel method for the automatic segmentation of polyp areas from colonoscopy images. Specifically, to accurately identify the polyp boundaries, we use a hybrid LGB color space in which the primary colors green and blue with the Lightness component of CIE-L×a×b are concatenated. . In this research work, Multi-Dimensional Cascades Network (MDCNet) is developed for the early stages of Colorectal Polyp Segmentation. It consists of two stages, stage 1 involves performing polyp location and rough segmentation using a probability-anatomical-prior-guided shallow-layer enhanced 3D location net. In order to lower the uncertain probabilities and false positive boundary points, a novel circular inference module and parameter Dice loss are also suggested. A multi-view 2.5D net made up of three 2D refinement subnetworks is used in stage 2 to thoroughly investigate the morphological features, making up for errors and missing spatial information of a single view. The proposed MDCNET-based segmentation model detects the early stage of polyps of colonoscopy images. The experimental results are validated using the Sensitivity, PPV, Dice similarity coefficient(DSC), and IOU values. The DSC and IOU values for the segmentation results of the proposed method are 0.9259, and 0.8938 in the Kvasir-SEG Data set and 0.8109, 0.8045 in the CVC-ClinicDB dataset respectively.

***Keywords:*** *Colorectal Cancer; Polyp Segmentation; Fully Convolutional Network; Cascading Mechanism.*

## 1. INTRODUCTION

Colorectal cancer (CRC) is one of the most common types of cancer in adults. It is the third most common cancer worldwide [1]. Although the overall colorectal cancer death rate has decreased in recent years, it has become more common in younger adults [2-3]. Therefore, it is important to prevent and treat colorectal cancer. The best way to reduce the risk of colorectal cancer is to have regular colon screening. Screening tests can find precancerous abnormalities so that they can be removed before they become cancer [3]. One type of colon abnormality is colon polyps. Usually, when colon cells become abnormal and grow out of control, clumps called polyps form. Some types of polyps can change into cancer over time. Therefore, early detection of polyps can increase the survival rate to 90%[4]. The most effective screening tool for detecting and evaluating polyps is colonoscopy. During a colonoscopy, the doctor inserts a thin, long, flexible tube into the patient's rectum. This instrument, named a colonoscopy, has a small video camera that allows the doctor to look at the entire inside of the colon. If polyps are found, they can usually be removed using a wire loop passed down the colonoscopy. Studies have shown that some polyps may not be detected at colonoscopy. In other words, the manual diagnosis of polyps is error-prone. This is for some reasons: First, The colonoscopy procedure is time-consuming. It usually takes approximately 20 to 30 minutes[2] . Second, some rare types of polyps are difficult to diagnose due to their flat nature, which requires extensive experience and the expertise of doctors. Therefore, automatic and accurate methods are required [5-6]. The use of automated methods, as an auxiliary diagnostic tool, can improve the speed and accuracy of doctors in diagnosing colorectal cancer and reduce misdiagnosis.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Medical image segmentation is one of the most common image-processing techniques for diagnosing, examining injuries, and locating lesions. This technique divides the medical image into several areas with the same characteristics [7]. Specifically, in polyp image segmentation, the poly be tissue is separated from natural colon tissues. As a result, useful information about the polyp's location, size, and boundaries is obtained. Although many algorithms have been proposed for automatic polyp image segmentation, diagnosing polyps from colonoscopy images is very challenging due to the extensive diversity in the polyp's size, color, texture, and shape. Figure 1 shows the challenge of polyp segmentation from a colonoscopy image. As seen in Fig. 1, the polyp's size can be small, medium, or large.

## 2. RELATED WORK

Over the past years, many efforts have been made to develop automatic medical image segmentation algorithms. Some primitive algorithms use thresholding techniques [8-9], edge detection[9-10], and region-growing methods [11-12]. These methods failed to achieve high accuracy due to the lesions' diverse and complex structure, especially in colonoscopy imaging. The polyp's tissue may resemble the normal colon tissue or the polyps may be seen flat on colonoscopic images. Therefore, more robust methods are required to segment the polyp images. Other approaches based on watershed algorithms, clustering techniques, and machine learning have shown better results for image segmentation. However, these approaches can only be used for limited tasks due to noise and lack of consistency in source data acquisition [13].

In recent years, machine learning and deep learning have been widely applied in medical imaging. Machine learning algorithms for polyp image detection mainly use traditional image processing techniques, requiring the manual design of feature extraction procedures and classifiers.Wang et al. [14] proposed a polyp detection framework including image preprocessing, feature extraction, feature selection, and classification steps. Zheng et al. [15] proposed a computer-aided diagnosis method based on image feature extraction and Fisher vector technology. Zhang et al. [16] proposed a polyp detection method based on SIFT features. These approaches rely on expertise and experience and have elevated data quality requirements.

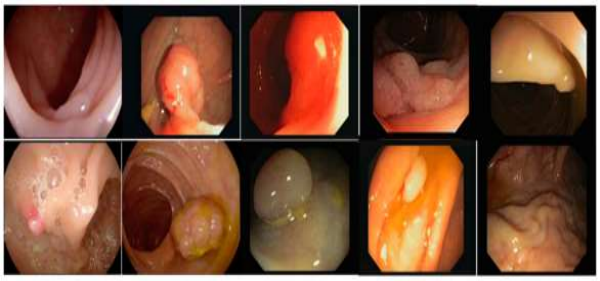

*Fig 1. Samples Of Colonoscopy Images With A Variety Of Polyp Conditions.*

In contrast, deep learning algorithms can automatically learn features and have relatively low data quality requirements. Zacharaki et al. [17] used support vector machines and neural networks to detect polyps in computed tomography colonoscopy, employing multiple texture feature extraction methods and comparing their effect. The final results showed that neural networks performed better in detecting polyps with regular textures and had better performance [18]. Early machine learning detection algorithms relied on the color and texture features of polyps, but the large color changes between polyps and the limited visibility of surface textures hindered the applicability of the algorithms [19]. Wang et al. [20] used a region-based CNN model to detect and classify polyps in colonoscopy images, which can detect potential polyp regions in images and classify them. Fang et al. [21] used a method based on convolutional neural networks (CNN) and region-based CNN (R-CNN), where CNN was used to generate candidate regions and R-CNN was used for classification and localization of these candidate regions, achieving polyp region localization in colonoscopy images.

Ronneberger et al. [22] developed the FCN architecture and presented the UNet architecture for biomedical image segmentation. U-Net consists of two main paths, the encoder, and the decoder. The encoder path is composed of stacking several convolutional layers and pooling layers. Convolution layers automatically extract features from inputs as feature maps, and pooling layers decrease the dimensions of the feature maps. On the other hand, the decoder increases spatial resolution by using up-sampling layers. Moreover, to improve segmentation accuracy, feature maps of the two paths are connected via a skip connection. Various modified versions of the U-Net architecture are presented for polyp image segmentation. For example, ResUNet is a well-known modified version of the U-Net architecture that replaces convolutional blocks in the U-Net with the residual

blocks. JHA et al. [23] used the ResUNet architecture to obtain the 0.5144 Dice similarity coefficient (DSC) in the polyp image segmentation task. The residual blocks using skip connections help to prevent the vanishing gradient problem in deep neural network training and provide faster convergence.

The ResUNet++ architecture, a modified version of the ResUNet, was introduced by JHA et al. [24] in another work. The output of the residual blocks in the ResUNet++ encoder passes through the squeeze-excitation module to give more importance to the necessary features. In addition, the ResUNet++ uses atrous spatial pyramid pooling (ASPP) blocks as a bridge between the encoder and the decoder to obtain useful multi-scale information for image segmentation tasks. Also, attention modules have been added to the ResUNet++ decoder, which receives encoder feature maps and decoder feature maps as input and determines which part of the feature needs more attention. They[24] achieved a DSC of 0.8133 for polyp image segmentation. This indicates that ResUNet performance improved due to the using squeeze excitation, ASPP, and attention blocks.

In [25], JHA et al. proposed other new architectures for polyp image segmentation and reported 0.8206 DSC on the Kvasir-SEG dataset. The best architectures they suggested are 1) UNet-Resnet34 architecture, which uses Resnet34 architecture as an encoder to extract features and achieves the best performance in terms of DSC, and 2) ColonSegNet architecture, which has fewer parameters than other architectures. Inspired by the UNet structure, Mahmud et al. [26] presented the deep PolypSegNet architecture, which includes depth dilated inception (DDI) modules, deep fusion skip (DFSM) modules, and deep reconstruction (DRM) module for polyp image segmentation. These modules integrate multi-scale feature maps generated at different levels of encoder and decoder[27]. Fan et al. [28] proposed the parallel reverse attention network (PraNet) to accurately determine polyp areas. PraNet uses a two-step process that involves predicting the initial regions of the polyp by parallel partial decoding (PPD), followed by the gradual refining of the polyp boundaries by a set of recurrent reverse attention (RA) modules. Tomar et al. [29] expanded the U-Net architecture using two decoders and attention blocks to automate polyp image segmentation and increase the generalizability of the network. Rauniyar et al. [30] provided an encoder-decoder architecture based on a dual-decoder attention network called DDANet and obtained a DSC of 0.8576 on the Kvasir-SEG dataset. Also, the DDANet model was tested on an unseen dataset with 200 images and achieved a DSC of 0.7874. From the quantitative results of the Kvasir-SEG and unseen test datasets, it can be concluded that the performance and generalizability of architectures are still low. Therefore, a more reliable method is needed to segment the polyp images.

The main goal of this paper is to employ a deep adversarial model based on a complete convolutional network to improve the segmentation performance of polys.. The segmentation network's end-to-end output may be realized by using the generator model based on the fully generated convolutional network. The discriminator is given the segmented image and the label image, and important properties of the learning data are strengthened by the adversarial training of the generator and discriminator, further improving the network's segmentation performance. A combination of the loss function and adversarial function is also proposed to further increase segmentation accuracy and relieve the poly-category imbalance problem. The rest of this paper is organized as follows: Section 3 describes the proposed methodology. Experiments and evaluation are explained in section 4. Finally, a conclusion is drawn in Section 5

## 3. PROPOSED METHODOLOGY OF POLYP SEGMENTATION

In this section, we provide a brief introduction to the main processing frameworks of this paper as follows: the full convolution structure for the generator and the discriminator structure are provided, and in addition, a combined loss function based on reducing category imbalance is then introduced.

### 3.1. Pre-Processing

Colonoscopy images of different patients have different dimensions. Therefore, in the first stage of preprocessing, the dimensions of all images are changed to 320×320. Then a new hybrid color space is created from the RGB and L×a×b color spaces. The lightness component is not separated from the color components in RGB color space. Therefore, the images are transferred from the RGB space to the CIE L × a × b space. Because the polyp tissue is similar to the normal colon tissue, the lightness channel of the L × a × b color space replaces the red channel. As a result, the hybrid LGB color space will be created. According to Figure 2, it can be

expected that the sharpness of the polyp edges will increase in the proposed hybrid color space. In the next step, in order to normalize the input, each color channel is normalized according to its maximum and minimum value. If 'x' is the desired color channel with the highest value of '$x_{max}$' and the lowest value '$x_{min}$', the normalized channel ('$x_n$' ) is calculated through the following Eqn, which has values between zero and one:

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}}.$$

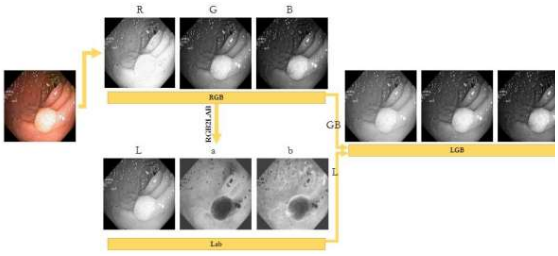


*Fig. 2. The Proposed Hybrid LGB Color Space.*

### 3.2 Multi-Dimensional Cascaded Network

The overall architecture of the proposed MDCNet is discussed in this section. As shown in Fig. 3, MDCNet is cascaded by a high-resolution(HR) multi-view 2.5D net and a Shallow-Layer (SL) Enhanced 3D location net. The original image V, which has the dimensions X, Y, and Z, is first down-sampled and delivered to stage 1. The shallow-Layer encoder-decoder module (SLEDM) is used in stage 1 to extract all of the semantic information. The circular Inference Module (CIM) receives both the extracted features and the likelihood anatomical prior. The general segmentation M is reached after a number of conclusions. Dice loss, a novel parameter, trains stage 1.

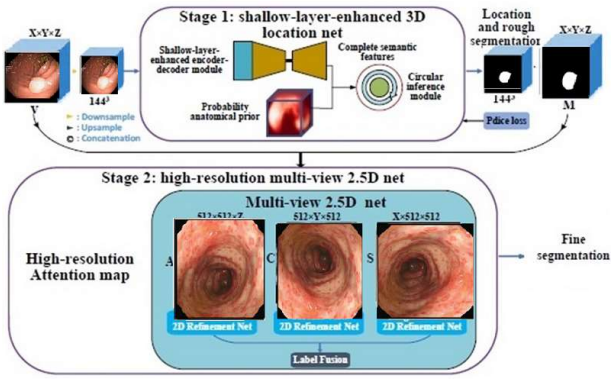


*Fig. 3: Over All Flow Diagram Of Mdcnet.*

In stage 2, a high-resolution attention map module that combines slicing, concatenation, and super-resolution is used to integrate V and M into 2D image slice sets from sagittal, coronal, and axial views with high image clarity and rich detail information. They are then given to a multi-view 2.5D net composed of three 2D refinement nets to perform precise segmentation. After this net's outputs are combined, a precise segmentation outcome is obtained.

#### 3.2.1 . Stage 1: Shallow-layer(SL)-enhanced 3D location net

Stage 1 places a focus on extracting comprehensive semantic features to learn about the position, organization, and shape of Polyp lesions, particularly small and asymmetrical lesions. Fig. 4 depicts the stage 1 construction.A type of asymmetric encoding-decoding structure, SLEDM is composed of numerous continuous multi-scale convolution blocks. Since shallow layers contain a lot of intricate details that are essential for accurate contour extraction and localization, we use a shallow-layer-enhanced method to appropriately increase their number of convolution blocks and channels during the encoding stage..

In the network's downsampling, the detail features frequently gradually disappear. However, the improved structure can improve the DL network's extracted detail features. The SLEDM operation can be described as follows. Assume that stage 1's input is $\mathbf{I} \in \mathbb{R}^{B,1,144,144,144}$ . The size of the variable is reflected in the high-dimensional space R. The batch size is represented by the first dimension for I, the channel number by the second dimension, and the image size by the final three dimensions. Following the encoding-decoding module, the data are is define as per Eqn(1)

$$\mathbf{D} = \mathbf{F(I)} \quad (1)$$

Where (•) denotes the module for enhanced shallow-layer encoding and decoding. In the multi-scale network, whose scales are $144^3$ and $72^3$, the first and second layers of the encoding stage are regarded as shallow layer. In this work, we increased the number of convolution blocks and channels in the shallow layer here. The number of convolution blocks is specifically specified to 3. The first layer's channel count steadily rises from 16 to 64, while the second layer's channel count stays the same at 64. Conv+bn+prelu make up the convolution blocks in (•). Each encoder and decoder block ends with the application of the residual module and dropout layer. The features that this module has extracted $\mathbf{D} \in \mathbb{R}^{\mathbf{B,32,144,144,144}}$

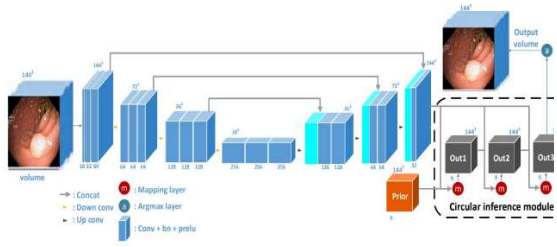


*Fig. 4: Detail Structure Of SL-Enhanced 3D Location Net.*

### 3.2.1.1 Probability anatomical prior

The segmentation of networks can be guided by anatomical information, eliminating minor location errors. Here, based on a multi-center training set, we suggest an anatomical likelihood prior. Assume that N examples with a size of 144×144×144 are included in the training set T RN,144,144,144, which stands in for T. 8 annotated Polyp lesions and the background are depicted by Lo, o [1, 9], which stands for the label categories. The i-th sample's x, y, and z-th voxel point can be written as

By eliminating minor location errors, anatomical information can direct the segmentation of networks. Here, based on a multi-center training set, we suggest a likelihood anatomical prior. Assume that the training set $\mathbf{P} \in \mathbb{R}^{\mathbf{N,144,144,144}}$ demonstrating **T** comprises **N** models with a size of 144×144×144. The label categories are represented by $\boldsymbol{L_o}$ , $\boldsymbol{o} \in [\mathbf{1}, \mathbf{9}]$ , indicating 8 annotated Polyp lesions and the background. The $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$-th voxel point of the $\boldsymbol{i}$-th sample can be expressed as per Eqn(2)

$$P(x,i,j,k) = L_0 \Rightarrow H(x,o,i,j,k) = 1, \quad (2)$$

As a result, the likelihood anatomical prior $P_{ar}$ is determined in the manner shown as per Eqn(3):

$$P(x,i,j,k) = L_0 \Rightarrow H(x,o,i,j,k) = 1$$
$$else\,if\ P(x,i,j,k) \neq L_0 \Rightarrow H(x,o,i,j,k) = 0$$
$$P_{ar} = \frac{\sum_{x=1}^{N} H(x,o,i,j,k)}{N} \quad (3)$$

Where $\mathbf{H} \in \mathbb{R}^{\mathbf{N,9,144,144,144}}$ is a binary score accumulator used to calculate $\mathbf{P}_{ar}$ , $\mathbf{P}_{ar} \in \mathbb{R}^{\mathbf{9,144,144,144}}$ . $P_{ar}(\boldsymbol{o}, i, j, k)$ denotes the probability of the i, j, k-th voxel point belonging to category $\boldsymbol{L_o}$. According to the computed probability anatomical prior, the more red a voxel point is, the more likely it is that it is associated with the corresponding lesion. The forecast is helped by the fact that we can roughly determine each lesion's location and shape for segmentation.

### .3.2.1.2 Circular inference module

In segmentation problems, the cores of the objects typically have high probabilities while the edges typically have uncertain probabilities. Here, we create a brand-new circular inference module to lower the hazy likelihood of blurred edges. Our circular inference module, which is comprised of three mapping layers , it is thought to function as a micro-attention mechanism. Each layer is predicated on the preceding output. The data are mapped to a 9-channel probability output between 0 and 1, which indicates the likelihood that each voxel corresponds to the appropriate organ. Each layer is made up of a 3D convolution (size of 111, stride of 1 and 9 output channels). In this case, an uncertain point is defined as a voxel with a probability between γ1 and γ2 is set as an uncertain point, where $0 < \gamma1 < \gamma2 < 1$ . We adjust the probability of uncertain locations in the following output by computing the weighted loss using manually annotated labels, the intermediate outputs of the circular truth, and the prior output(Li et al.,2022).

The following describes the corresponding process is defined in Eqn(4)

$$\begin{aligned} \mathbf{Out1} &= \mathbf{m_1}[\mathbf{cat(D,P)}] \\ \mathbf{Out2} &= \mathbf{m_2}[\mathbf{cat(D,Out1)}] \\ \mathbf{Out3} &= \mathbf{m_3}[\mathbf{cat(D,Out2)}] \end{aligned} \quad (4)$$

Where $\{(\bullet), \boldsymbol{i} \in [\mathbf{1}, \mathbf{3}]\}$ represent mapping layers, cat is concatenation, and $\{\mathbf{Out}\boldsymbol{i}, \boldsymbol{i} \in [\mathbf{1}, \mathbf{3}]\}$

$\in \mathbb{R}^{B,9,144,144,144}$ is the intermediate output. For the following prediction, which is detailed below, we calculate the weighted loss function between the manually annotated label and the prior outputs.

$$\mathbf{Loss} = \frac{\sum_{i=1}^{3} W_i L_i[\mathbf{label}, \mathbf{Out}i]}{3}$$

where $L_i[\mathbf{label}, \mathbf{Out}i]$ represents the loss between **label** and **Out**$i$. $\{W_i, i \in [1, 3]\}$ represents the weights between different losses, which are set to 0.5, 1, and 1.5 in this work.

Therefore, the output volume $O$ can be defined as per Eqn(5):

$$O = \mathbf{argma}(\mathbf{Out3}, \mathbf{2}) \quad (5)$$

where **argmax**( • , $i$) represents the argmax function along the $i$-th axis, $O \in \mathbb{R}^{B,144,144,144}$.

The module lowers the boundary's uncertain probability over numerous repetitions. The number of iterations can change depending on the task at hand and the environment.

### 3.3 . Multi-view 2.5D net for Polyp segmentation

Stage 1 identifies and roughly segments the Polyp lesions, directing stage 2's segmentation. However, while downsampling in the 3D structure, certain fine details are lost. Stage 2 includes a high-resolution attention map(HRAM) module and a multi-view 2.5D net and is built on a 2D structure with reduced image volume and memory usage. The result of stage 1 and the original image serves as the input for stage 2 of the HRAM, which reconstructs three HRs 2D image slices with the letters S, C, and A. The features lost in stage 1 are captured in the HR image input, boosting the edge information of the Polyp lesions, we concatenate and upsample the image slices to produce three HR 2D slice sets by separating the 3D V and M from stage 1 along the sagittal, coronal, and axial views. We have {Si , **C**j , **A**k | $i \in [1, X]$, $j \in [1, Y]$, $k \in [1, Z]$}. All of the image slices in these three sets are 512 by 512 pixels in size. In this system, the different levels of features are fusion for BT segmentation.

### 3.4. Loss Function

. In this paper, the absolute error loss $L1$ and the minimization of the perceptual loss $Lp$ are thoroughly examined, and a loss function appropriate for Colonoscopy images is developed. As we all know, norm loss functions $L1$ and $L2$ will have fuzzy problems regarding image segmentation. The loss function $L_1$ , the minimum absolute deviation function, has higher robustness and less fuzziness than the loss function $L_2$. Therefore, we introduce $L_1$ to encourage the network to segment an image with higher similarity to the ground truth image by measuring the pixel-level value difference between them. The loss function is as follows:

$$L_1 = \mathcal{L}\left(\hat{I}, I^*\right) = \sqrt{\left\|\hat{I} - I^*\right\|^2 + \varepsilon^2} \quad (6)$$

where $\hat{I}$ and $I*$ represent segmented images and ground truth images from the training dataset, respectively; $\varepsilon$ represents the error offset. The perceptual loss function is an advanced feature from the $block5$ layer of the pre-trained Cascaded network.

The perceptual loss function is primarily calculated by comparing the features taken from the image to a cascaded network that has already been trained. The following is what we propose:

$$L_p = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left[\varphi\left(\hat{I}\right)_{i,j} - \varphi\left(I^*\right)_{i,j}\right]^2 \quad (7)$$

where $\varphi$ represents the feature map of the pre-trained cascaded network. $H$ and $W$ are the height and width of the feature map. The total loss function is expressed as:

$$L = \lambda_1 L_1 + \lambda_2 L_p \quad (8)$$

where $\lambda 1$ and $\lambda 2$ represent the weight of the absolute error loss function and the perceptual loss function in the total loss function, respectively.

## 4. EXPERIMENTS AND EVALUATION

### 4.1. DATASET

The proposed method is evaluated on three open-access datasets such as ETIS [36], CVC-ClinicDB [34] and Kvasir-SEG [35]. The ETIS Data set contains 196 images collected from 34 colonoscopy videos. The image size is 1225 × 966. The Kvasir-SEG dataset contains thousands of diverse images of the small, medium, and large polyps with ground truth images created manually by gastroenterologists from a hospital in Norway. The resolution of the images in this dataset varies from 482×332 to 1072×1920. This dataset is used to train and validate the proposed X cep-MResUNet model because it contains more images and a variety of polyps than other datasets. For this purpose, 80% of data are randomly selected as training dataset, 10% as validation dataset, and the rest as testing dataset. The random selection procedure is repeated five times, and the average evaluation criteria values are reported. Besides, the generalizability capability of the proposed architecture is checked on the CVC-ClinicDB dataset. The CVC-ClinicDB dataset contains 612 polyp colonoscopy images with ground truth segmentation masks. The dimensions of all images are equal to 288×384.

### 4.2. Evaluation Metrics

Several popular metrics are adopted to evaluate the performances, including the Precision, Specificity, Accuracy. Dice Similarity Coefficient (DSC) , Intersection over Union (IoU) . It is given in the following Eqns

$$\mathrm{Pr} = \frac{TP}{TP + FP}$$

$$Sp = \frac{TN}{TN + FP}$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}$$

$$F1 = 2.\frac{Se.\mathrm{Pr}}{Se + \mathrm{Pr}}$$

Where T P, TN, FP, and FN denote true positive, true negative, false positive and false negative, respectively.

$$\mathrm{IoU} = \frac{A \cap B}{A \cup B},$$

$$\mathrm{DSC} = \frac{2|A \cap B|}{|A| + |B|},$$

where A and B represent pixel sets for the ground truths and their detection results, respectively

### 4.3 EXPERIMENTAL RESULTS

The experimental results of the proposed MDCNET method using in Kvasir-SEG dataset are given in Table 1 and the CVC-ClinicDB dataset is given in Table 2.

*Table 1. Comparison Of The Different Polyp*

| Method | IoU | DSC | Acc | Se | Sp | Pr |
|---|---|---|---|---|---|---|
| UNet [22] | 0.93 | 0.96 | 0.99 | 0.96 | 0.99 | 0.9657 |
| ResUNet [33] | 0.91 | 0.95 | 0.98 | 0.96 | 0.99 | 0.955 |
| ColonSegNet [25]] | 0.93 | 0.96 | 0.99 | 0.97 | 0.99 | 0.9601 |
| ResUNet++ [24] | 0.93 | 0.96 | 0.9675 | 0.97 | 0.9545 | 0.9656 |
| PolypSegNet [26] | 0.90 | 0.94 | 0.92 | 0.94 | 0.9124 | 0.9434 |
| Hardnet-mseg [27] | 0.86 | 0.92 | 0.98 | 0.97 | 0.99 | 0.8926 |
| Pranet [28] | 0.91 | 0.94 | 0.95 | 0.96 | 0.9423 | 0.9543 |
| **Proposed MDCNET** | **0.94** | **0.97** | **0.99** | **0.97** | **0.99** | **0.9691** |

*Segmentation Methods In The Kvasir-SEG Dataset.*

Comparing the proposed method with other modified ResUNet architectures [22-24,33] in Table 1 shows that the proposed method performs well on the Kvasir-SE test dataset. This is because there is a relatively large gap between the quantitative criteria obtained by the architectures designed in [22-24,33]. However, the Hardnet-mseg [27] and PolypSegNet [26] architectures have better performance than the proposed Xcep-MResUNet, as the formers have filters of different sizes that can retrieve spatial information more accurately. On the other hand, as reported in Table 2, the proposed architecture achieves good performance compared to other ResUNet-based architectures due to the use of the transfer learning technique.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 2. Comparison Of The Different Polyp Segmentation Methods In The CVC-Clinicdb Dataset.*

| Method | IoU | DSC | Acc | Se | Sp | Pr |
|---|---|---|---|---|---|---|
| UNet [22] | 0.9111 | 0.9507 | 0.9464 | 0.9353 | 0.9354 | 0.9439 |
| ResUNet [33] | 0.8727 | 0.9248 | 0.9801 | 0.9044 | 0.9939 | 0.9559 |
| ColonSegNet [25]] | 0.8980 | 0.9444 | 0.9833 | 0.9483 | 0.9900 | 0.9455 |
| ResUNet++ [24] | 0.9290 | 0.9354 | 0.9439 | 0.9354 | 0.9439 | 0.9354 |
| PolypSegNet [26] | 0.8890 | 0.9400 | 0.9878 | 0.9840 | 0.9654 | 0.9439 |
| Hardnet-mseg [27] | 0.8873 | 0.9392 | 97.21 | 0.9895 | 0.9354 | 0.9439 |
| Pranet [28] | 0.9290 | 0.9354 | 0.9439 | 0.9700 | 0.9100 | 0.9567 |
| **Proposed MDCNET** | **0.9351** | **0.9651** | **0.9895** | **0.9711** | 0.9944 | 0.9612 |

In addition to quantitative evaluations, a visual review of the results helps to understand the performance of the proposed method. Therefore, samples of colonoscopic images from Kvasir-SEG and CVC-ClinicDB datasets are shown in Figures (5) and (6), respectively, along with the segmentation results of the proposed method. As can be seen in the figures, the proposed architecture has successfully detected different polyps; however, there are still challenges in the case of polyps hidden behind colon wrinkles.

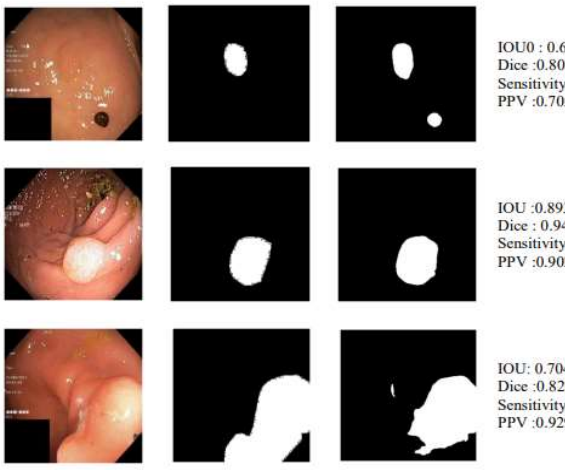


*Fig.5. Visual results from proposed MDCNET architecture on the Kvasir dataset. Each subfigure, from left to right shows a colonoscopy image, ground truth, and the segmentation result of the proposed architecture.*

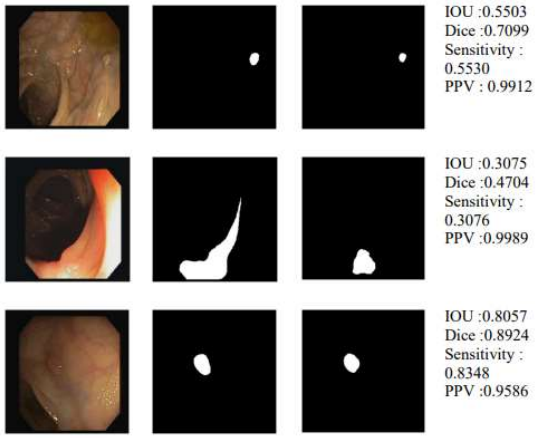


*Fig.6. Visual results from proposed MDCNET architecture on the CVC-ClinicDB dataset. Each subfigure, from left to right shows a colonoscopy image, ground truth, and the segmentation result of the proposed architecture.*

## 5. CONCLUSION

In this paper, an efficient method is presented for polyp image segmentation in colonoscopy images. First, image values transfer from RGB color space to the proposed hybrid LGB color space. Then, partial preprocessing was performed to equalize the dimensions of the images and normalize them. To determine the polyp area from the colonoscopy image, the proposed MDCNET is used. The primary contribution of MDCNET lies in its sub-network cascading process, which enhances network robustness while providing better-contextualized features for underwater image improvement. The proposed MDCNET is extensively evaluated on three poly image segmentation benchmarks from the ETIS, Kvasir-SEG dataset, and CVC-ClinicDB datasets. Experiment results show that MDCNET outperforms its baseline by a wide margin and achieves performance comparable to standard poly segmentation methods. Future work on the task of polyp image segmentation can provide a more diverse dataset or can also extend this study to other segmentation tasks.

## REFERENCES

[1]. J. Asplund, J. H. Kauppila, F. Mattsson, and J. Lagergren, ``Survival trends in gastric adenocarcinoma: A population-based study in Sweden," Ann. Surgical Oncol., vol. 25, no. 9, pp. 2693-2702, Sep. 2018.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[2]. Siegel, R. L. et al. Colorectal cancer statistics, 2020. CA: A Cancer Journal for Clinicians 70, 145–164 .2020.

[3]. Franz M, Scholz M, Henze I, Röckl S, Gomez LI. Detection of colon polyps by a novel, polymer pattern-based full blood test. Journal of Translational Medicine. 2013;11(1):1-9.

[4]. Yang K, Chang S, Tian Z, Gao C, Du Y, Zhang X, et al. Automatic polyp detection and segmentation using shuffle efficient channel attention network. Alexandria Engineering Journal. 2022;61(1):917-26.

[5]. Stoitsis, J. et al. Computer aided diagnosis based on medical image processing and artificial intelligence methods. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 569, 591–595 .2006.

[6]. Riegler, M., Lux, M., Griwodz, C., Spampinato, C., de Lange, T., Eskeland, S.L., Pogorelov, K., Tavanapong, W., Schmidt, P.T., Gurrin, C., Johansen, D., Johansen, H., Halvorsen, P.: Multimedia and Medicine: Teammates for Better Disease Detection and Survival. In: Proceedings of ACM Multimedia (ACM MM). pp. 968{977. ACM (2016)

[7]. Pham, D.L., Xu, C., Prince, J.L.: Current methods in medical image segmentation. Annual review of biomedical engineering 2(1), 315{337 .2000.

[8]. Ilhan, U. Ilhan, A., Polyp lesion segmentation based on a new threshold approach. Procedia computer science. 2017;120: 580-587.

[9]. Gupta, N., P. Bhatele, Khanna, P., Glioma detection on Polyp MRIs using texture and morphological features with ensemble learning. Biomedical Signal Processing and Control. 2019; 47: 115-125.

[10]. Todd, C., Kirillov, M., Tarabichi, M., Naghdy, F. & Naghdy, G. An analysis of medical image processing methods for segmentation of the inner ear. In Proceedings of the IADIS International Conference Computer Graphics, Visualization, Computer Vision and Image Processing 2009, CGVCVIP 2009. Part of the IADIS MCCSIS 2009 (2009).

[11]. Rahima, Z., et al, Segmentation of low-grade gliomas based on the growing region and level sets techniques. in 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). IEEE. 2018.

[12]. Li, Q., Gao, Z., Wang, Q., Glioma segmentation with a unied algorithm in multimodal MRI images, IEEE Access. 2018; 6: 9543- 9553.

[13]. Haque IRI, Neubert J. Deep learning approaches to biomedical image segmentation. Informatics in Medicine Unlocked. 2020;18:100297.

[14] Wang, Y., Dorner, S., & Ecker, R. (2010). A framework for automatic polyp detection in colonoscopy images. Medical image analysis, 14(4), 616-629.

[15] Zheng, Y., Wang, X., Song, Y., & Wang, H. (2018). Computer-aided diagnosis for colonoscopy by using bag-of-visual-words and Fisher vector techniques. Journal of medical systems, 42(2), 31.

[16] Zhang, X., Chen, Y., & Song, Y. (2016). A novel approach for automated polyp detection in colonoscopy images via SIFT features. Journal of medical systems, 40(6), 136.

[17] Jayachandran, A & Dhanasekaran, R, "Severity Analysis of Brain Tumor in MRI Images using Modified Multi-Texton Structure Descriptor and Kernel- SVM, The Arabian Journal of science and engineering October 2014, Volume 39, Issue 10, pp 7073-7086,(2014).

[18] A comparative study of texture features for the detection of colonic polyps in computed tomography colonography", Zacharaki et al. (2009)

[19] Tajbakhsh, Nima, et al. "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?" IEEE Transactions on Medical Imaging, vol. 35, no. 5, Mar. 2016, pp. 1299–312, https://doi.org/10.1109/tmi.2016.2535302.

[20] Wang, P., Xiao, X., Glissen Brown, J. R., & Berzin, T. M. (2018). Automatic detection of colonic polyps in endoscopic images using region-based convolutional neural networks. IEEE Journal of Biomedical and Health Informatics, 22(5), 1495-1505.

[21] Fang, Y., Zhang, J., Zhang, Y., & Gao, Y. (2016). Polyp Detection Using Convolutional Neural Networks and Region-Based Fully Convolutional Networks. International Conference on Medical Image Computing and Computer-Assisted Intervention, 9902, 62-70.

[22]. Ronneberger, O., Fischer, P., Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. 2015; 234-241.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[23]. Jha D, Smedsrud PH, Riegler MA, Halvorsen P, Lange Td, Johansen D, et al., editors. Kvasir-seg: A segmented polyp dataset. International Conference on Multimedia Modeling; 2020: Springer.

[24]. Jha D, Smedsrud PH, Riegler MA, Johansen D, De Lange T, Halvorsen P, et al., editors. Resunet++: An advanced architecture for medical image segmentation. 2019 IEEE International Symposium on Multimedia (ISM); 2019: IEEE.

[25]. Jha D, Ali S, Tomar NK, Johansen HD, Johansen D, Rittscher J, et al. Real-time polyp detection, localization and segmentation in colonoscopy using deep learning. Ieee Access. 2021;9:40496-510.

[26]. Mahmud T, Paul B, Fattah SA. PolypSegNet: A modified encoder-decoder architecture for automated polyp segmentation from colonoscopy images. Computers in Biology and Medicine. 2021;128:104119.

[27]. Huang C-H, Wu H-Y, Lin Y-L. Hardnet-mseg: a simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps. arXiv preprint arXiv:210107172. 2021.

[28]. Fan D-P, Ji G-P, Zhou T, Chen G, Fu H, Shen J, et al., editors. Pranet: Parallel reverse attention network for polyp segmentation. International conference on medical image computing and computer-assisted intervention; 2020: Springer.

[29]. Jesu Prabhu A and Jayachandran, A, "Mixture Model Segmentation System for Parasagittal Meningioma Brain Tumor Classification based on Hybrid Feature Vector" Journal of Medical System, vol 42, issues 12, 2018.

[30]. Jayachandran, A and R.Dhanasekaran ,(2017) 'Multi Class Brain Tumor Classification of MRI Images using Hybrid Structure Descriptor and Fuzzy Logic Based RBF Kernel SVM' , Iranian Journal of Fuzzy system , Volume 14, Issue 3, pp 41-54 , 2017.

[31].Chen, T., Wang, N., Wang, R., Zhao, H., Zhang, G., 2021. One-stage cnn detector-based benthonic organisms detection with limited training dataset. Neural Networks 144, 247–259.

[32]. Jayachandran, A, 'Abnormality segmentation and Classification of multi model brain tumor in MR images using Fuzzy based hybrid kernel SVM' International Journal of Fuzzy system , published by Springer, Volume 17, Issue 3, pp 434-443,2018..

[33].Zhang, J., Lv, X., Zhang, H.: AResU-Net: Attention residual U-net for Polyp lesion segmentation. Symmetry 12(5), 721 (2020).

[34]. J. Bernal, F.J. S´anchez, G. Fern´andez-Esparrach, D. Gil, C. Rodríguez, F. Vilari˜no, Wmdova maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians, Comput. Med. Imag. Graph. 43 (2015) 99–111.

[35]. Jha D, Smedsrud PH, Riegler MA, Halvorsen P, Lange Td, Johansen D, et al., editors. Kvasir-seg: A segmented polyp dataset. International Conference on Multimedia Modeling; 2020: Springer.

[36]. Mahiba C, A Jayachandran,"Severity analysis of diabetic retinopathy in retinal images using hybrid structure descriptor and modified CNNs",Measurement,Vol 135,PP 762-767,2019.

[37]. Milletari F, N. Navab, and S. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in Proc.- Int. Conf. 3D Vis., pp. 565–571, 2016.