



**HAL**  
open science

# Segmentation of Polyps in Gastrointestinal Tract Images

Sabrina Nasrin, Javaneh Alavi, Pamila Viswanathan

► **To cite this version:**

Sabrina Nasrin, Javaneh Alavi, Pamila Viswanathan. Segmentation of Polyps in Gastrointestinal Tract Images. [Research Report] University of Alberta. 2021. hal-03501140

**HAL Id: hal-03501140**

**<https://hal.science/hal-03501140>**

Submitted on 23 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Segmentation of Polyps in Gastrointestinal (GI) Tract Images

Sabrina Nasrin

*dept. of Computing Science*

*University of Alberta*

Edmonton, Alberta-Canada

snasrin@ualberta.ca

Javaneh Alavi

*dept. of Computing Science*

*University of Alberta*

Edmonton, Alberta-Canada

Javaneh@ualberta.ca

Pamila Tina Viswanathan

*dept. of Computing Science*

*University of Alberta*

Edmonton, Alberta-Canada

pviswana@ualberta.ca

**Abstract**—Detecting abnormal tissues can be overlooked during body screening procedures including endoscopy, bronchoscopy, and colonoscopy. Colonoscopy is a routine screening procedure that can examine inside of the large intestine. However, observants might not be able to detect anomalies at initial phase. Therefore, a precise method is needed to detect the abnormalities. In this paper, we have implemented three different convolutional neural networks to segment polyps in gastrointestinal tract images. First, UNet which consist of two parts contraction and expansion for segmenting medical images. In this model data augmentation is performed with elastic deformations to yield accurate results with very few annotated images. Then, we implemented TriUnet which consists of three UNet models. The last model DivergentNets is an ensemble of five segmentation models named as TriUnet, Unetplusplus, FPN, DeeplabV3 and DeeplabV3plus. We have also tested images by using color correction, image pyramid and specularly removal. Our results suggest that when we combine different segmentation models as DivergentNets, it produces better results than UNet and TriUnet.

## I. INTRODUCTION

Colorectal cancer is one of the most common malignancy of the digestive tract and cause of cancer-related death in both sexes [1]. Polyps, are abnormal growth tissue from the mucous membrane (Fig. 1). Colorectal polyps are usually benign while some might be precancerous or even cancerous over a period of 5 to 15 years [2]. Colorectal polyps may be lethal at their later stages. Therefore, it is critical to detect and remove them at earlier stage.

Colonoscopy is a safe diagnostic procedure approach to detect polyps. However, previous studies showed that polyps can sometimes be unnoticed during colonoscopy procedure. Moreover, manually detecting polyps needs a lot of time and efforts as well as it is susceptible to physicians' bias. Therefore, it is required to find an automated and robust approach to detect them. Finding a way to spot polyps automatically can assist pathologist to increase chance of detecting polyps at early stage of their formation, which is crucial for prevention and survival from colorectal cancer [3].

Developments in computer vision and artificial intelligent algorithms has been influenced different fields of science and industry. It also has been simplifying the most complicated problems embarked in wide variety of professions including engineering, art, and medicine. The trend shows the increasing demand on usage of automation for verity of tasks in a real operation. One of the interesting area is image segmentation. Segmentation has been widely utilized in different fields such as robotics to find and recognize objects in environment [4], smart cities for urban surveillance, road and building segmentation [5], self driving cars [6], medical images [7], and many other areas. Segmentation is an important stage of the image recognition system.

The idea behind image segmentation is to assign same label to pixels that have the same specific characteristics. Image segmentation

and object detection techniques can divide image to meaningful parts that are easier to analyze and interpret [3]. Image segmentation techniques can assist detecting abnormalities in medical images. These techniques can assist pathologist in many ways such as monitoring and improving the diagnostic ability. In this paper, we have experimented three different architectures, UNet, TriUnet, and DivergentNets for polyp segmentation to decrease the chance of polyps to be overlooked.

The paper is organized as follows. Section II describes the related work in the literature. Section III describes the data set. In Section IV we discussed the Implemented methods while Section V reports the results of the implementation. Section VI is the discussion about the implementation results. Finally, Section VII draws some conclusion and outline some future works.

## II. LITERATURE REVIEW

Image segmentation being one of the most important domains in computer vision starting from early 1970s with study involving region growing techniques and optimization approaches. As colonoscopy screening suffers from human errors and failure to fully recognize polyps, computer-aided detection, localization, and segmentation methods can help to improve colonoscopy procedures. Among them the task of polyp segmentation attracted much in recent years.

The early approaches came under shape-based segmentation category where predefined polyp shape models were used. In [8], they used ellipse fitting techniques based on image curvature, edge distance, and intensity values for polyp detection. In [9], they implemented the Canny edge detector to process prior-filtered images, identifying the relevant edges using a template matching technique for polyp segmentation. Recently an increasing trend of colonoscopy image analysis have become more automated and integrated by making use of the latest machine learning techniques.

The paper [10] proposes a novel ResUNet++ architecture for medical image segmentation. A deep learning architecture for segmentation of biomedical applications is UNet [11]. The road image extraction has been provided by the ResUNet [12] architecture, which acts as the basis for the proposed architecture. They have created a new Kvasir-SEG dataset by annotating the existing Kvasir dataset. The ResUNet++ consists of one stem block followed by three encoders, Atrous Spatial Pyramidal Pooling (ASPP) and three decoder blocks. The Residual unit consists of a combination of normalization, Rectified Linear Unit (ReLU) activation and convolution layers. It helps to improve the design of the neural network by decreasing the parameters helps to propagate information within the network without degradation and increasing performance in segmentation. The encoder consists of a convolution block including a normalization layer and an identity mapping that connects the input and output of the encoder. The output of the encoder is sent to squeeze and excitation block. This block increases the sensitivity of the relevant features and suppress the unnecessary features. The ASPP enlarges the field-of-view of the filters by resampling features at multiple scales. This this architecture it also acts as a bridge between the encoder and the decoder. The decoder also consists of residual units where attention block increases the effectiveness, nearest neighbour upsampling from lower levels and concatenation with feature maps from the corresponding encoder. It is passed through ASPP and a 1x1 convolution layer that provides the segmentation map. Evaluation was done using two datasets with all the methods, it has been proven that ResUNet++ architecture outperforms the UNet and ResUNet in terms of accurate predictions.

In [13], they presented a new strategy for the delineation of gastrointestinal polyps from endoscopic images based on a direct extension of common encoder-decoder networks for semantic segmentation. In this approach, they

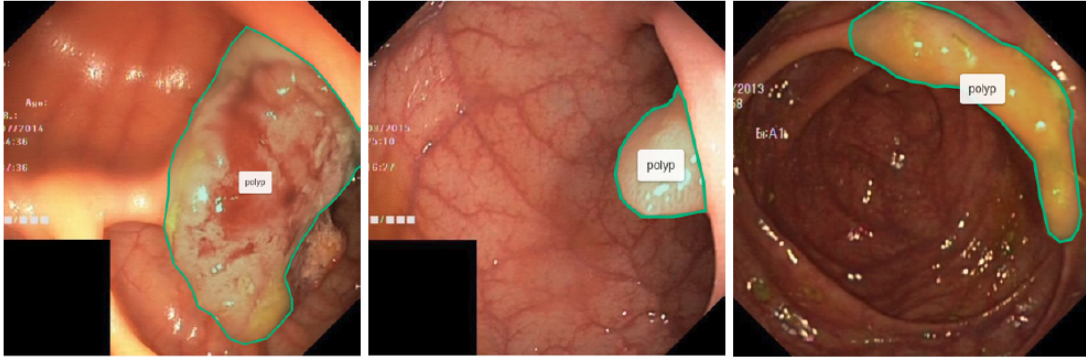


Fig. 1. Example frames from the Kvasir dataset where polyp tissue are shown in green [3]

proposed two sequentially stacked pretrained encoder-decoder networks where the second network takes the output of the first network as input. The initial prediction generated by the first network acts as an attention mechanism enabling the second network to focus on interesting areas within the image, thereby improving the quality of its predictions. In [14], they developed a framework of polyp image segmentation by a deep learning approach, specially a convolutional neural network. This framework is based on image preprocessing and two types of SegNet architecture to obtain the segmented polyp image. The paper has demonstrated state-of-the-art performance on both VGG-16, and VGG-19 networks for training and testing data to address colorectal cancer screening rates. In [15], they describe a polyp segmentation algorithm that was developed based on fully convolutional network models. In [3], they demonstrated the use of their dataset with two segmentation approaches. The method is based on the efficient FCM [16] unsupervised clustering architecture and the second one is based on the deep-learning ResUNet [17] architecture and utilized the advantage of the residual block. In [18], they proposed ColonSegNet for real-time polyp identification and delineations that can potentially transform current clinical practices and minimise miss-detection rates.

As clinicians miss polyps between 14% to 30%

during standard screenings of gastrointestinal tract, the paper [19] suggested a novel approach to improve segmentation performance by introducing a grid segmentation framework. They have applied the framework to U-Net and Pix2Pix architectures called, Grid-U-Net and Grid-GAN, respectively. Their hypothesis was that these receptive fields of U-Net can correspond to the grid cells which would correspond to a segmentation precision when the grid cells are classified. The same grid framework also applied to the Pix2Pix model for comparison. The grid segmentation networks have shown particularly promising results for lower precision segmentations, which produced better results than ordinary pixel-to-pixel segmentation at the cost of less precision. In [20] authors mention three different techniques for segmentation, classical image processing, extracting features, and using classifiers for segmentation and convolution neural networks.

### III. DATASET

There are several datasets available for polyp images that consist of ground truth and corresponding segmentation mask, including ASU-Mayo, CVC-Colon DB, ETIS-Larib Polyp DB, CVCClinicDB, Endoscene, Kvasir, and Kvasir-SEG Dataset.

#### A. CVC-612

CVC-ClinicDB is an open-access dataset of 612 images with a resolution of 384×288 from

31 colonoscopy sequences. It is used for medical image segmentation, in particular polyp detection in colonoscopy videos [21].

### B. PolypGen

PolypGen dataset is curated from 6 unique centres incorporating more than 300 patients. The dataset includes both single frame and sequence data with 3446 annotated polyp labels with precise delineation of polyp boundaries verified by six senior gastroenterologists [22].

### C. Kvasir-Seg

The Kvasir-Seg dataset contains 1000 polyp images and their corresponding ground truth from the Kvasir Dataset v2 [3].

## IV. IMPLEMENTED METHODS

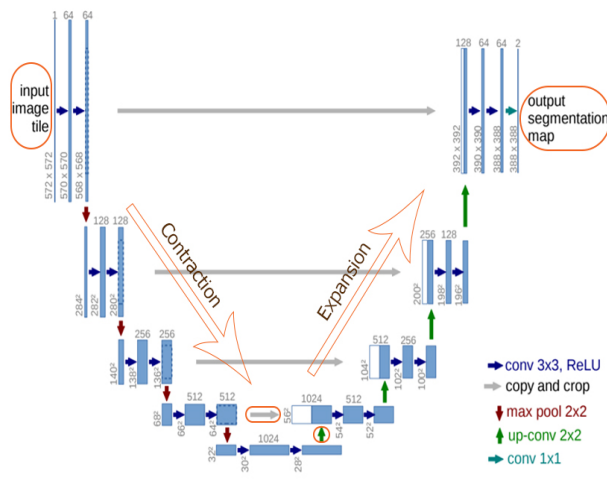


Fig. 2. UNet Architecture [11]

### A. UNet

Olaf Ronneberger developed UNet architecture, a fully convolutional neural network for biomedical image segmentation. The Unet model has been implemented using TensorFlow 2.0. this model is designed to work with very few images to yield a more precise segmentation. The main modification is implemented in the upsampling part by adding more feature channels. Since this network does not have

fully connected layers, it uses the parts of the segmentation for which the full context is available. As this network has very limited training data available, it is trained to learn invariance to deformations.

Its architecture consists of mainly two parts, contraction (left side) and expansion (right side). The left side is the same as that of the typical convolution network. For each down-sampling, the number of feature channels is doubled. On the expansion side, other than the upsampling of the feature map the corresponding feature map is concatenated. A total of 23 convolution layers are used in this network.

Caffe model [23] is used to train images and their segmentation maps. The energy is computed by pixel-wise soft-max (Eqn. 4) rest over cross-entropy function (Eqn.2). When the value of softmax is 1, indicates the maximum activation and 0 for the rest.

$$p_k(x) = \exp(a_k(x)) / \left( \sum_{k'=1}^K \exp(a_{k'}(x)) \right) \quad (1)$$

$p_k(x)$  denotes approximate maximum function

$a_k(x)$  denotes activation in feature channel  $k$ .  $k$  is the number of classes

$$E = \sum_{x \in \Omega} \omega(x) \log(p_{\ell(x)}(x)) \quad (2)$$

$\ell : \Omega \rightarrow 1, \dots, K$  denotes true label of every pixel

$\omega : \Omega \rightarrow \mathbb{R}$  denotes weighted map

Morphological operations are used to separate small borders. This process is done before training by using weighted maps (Eqn. 8) for each ground truth segmented so that the network learns to separate thin borders between segments. The initial weights are drawn from a Gaussian distribution.

$$\omega(x) = \omega_c(x) + \omega_0 \cdot \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2}\right) \quad (3)$$

$d_1 : \Omega \rightarrow \mathbb{R}$  denotes distance to border of nearest cell

$d_2 : \Omega \rightarrow \mathbb{R}$  denotes distance to border of second nearest cell [11].

### B. TriUNet

TriUNet is a combination of three UNets [11] placed in a triangular shape as depicted in Fig. 3. TriUNet takes a single image as an input which is then passed to two separate UNets in parallel with different randomized weights. These two UNets produce two feature vectors. These two outputs are concatenated before passing it to the third UNet as an input. The third UNet produces the final prediction mask. The loss is estimated by back-propagating through the entire network using the dice coefficient of the mask that corresponds to the main class and the ground truth. From the architecture of TriUNet it is clearly seen that different UNet learn different interpretations of the data, which are subsequently integrated into a single final output [24].

### C. DivergentNets

DivergentNets consists five high performance segmentation models such as UNet++, DeeplabV3, DeeplabV3+, FPN and TriUNet as illustrated in Fig. 4. Models were selected on the basis of their performance in segmentation tasks. DivergentNets is constructed from the idea that cumulative decisions taken from multiple high performance model can give more precise decision than the predictions from a single model. The models are trained separately before with N number of epochs to get the best checkpoints. The best checkpoint of each model is selected to be combined in DivergentNets. Like the TriUNet, DivergentNets also take a single image as an input which is passed through all the five models separately. Here, masks for two classes are predicted, background and polyp, with no overlap between them. The masks produced by each intermediate model represent the divergent views on

the data. The pixels between the output masks from each model is averaged and rounded to the nearest integer (either 0 or 1) to produce the final prediction [24].

### D. Single-channel Dice

The loss function is named as single-channel dice loss as it only considers one channel when calculating error. The equation is shown below.

$$Dice = \frac{2 \cdot |A_n \cap B_n|}{2 \cdot |A_n \cap B_n| + |B_n \setminus A_n| + |A_n \setminus B_n|} \quad (4)$$

where n represents the class for which loss should be calculated for. In this case, we only calculate loss for the polyp class and ignore the background [24].

### E. Tested Techniques

#### 1) Color Correction

To fix the color distortion, color correction was implemented on input images so that medical images from different sources will have standard color and will be more appropriate for segmenting the polyps. On implementing this technique, the results showed some improvements.

#### 2) Specularity Removal

Specular highlights are due to the reflection of the light shone by the colonoscopy torch onto the device's camera. Consequently, this generates areas of the image with high intensity, causing the loss of their information [25]. Adding the specularity removal in our case did not give good results as they smoothen the image and sometimes the polyps also are smoothened.

### F. Evaluation Metrics

#### 1) Accuracy

In general, the accuracy metric measures the ratio of correct predictions over the total number of instances evaluated [26].

$$Accuracy(acc) = \frac{tp + tn}{tp + fp + tn + fn} \quad (5)$$

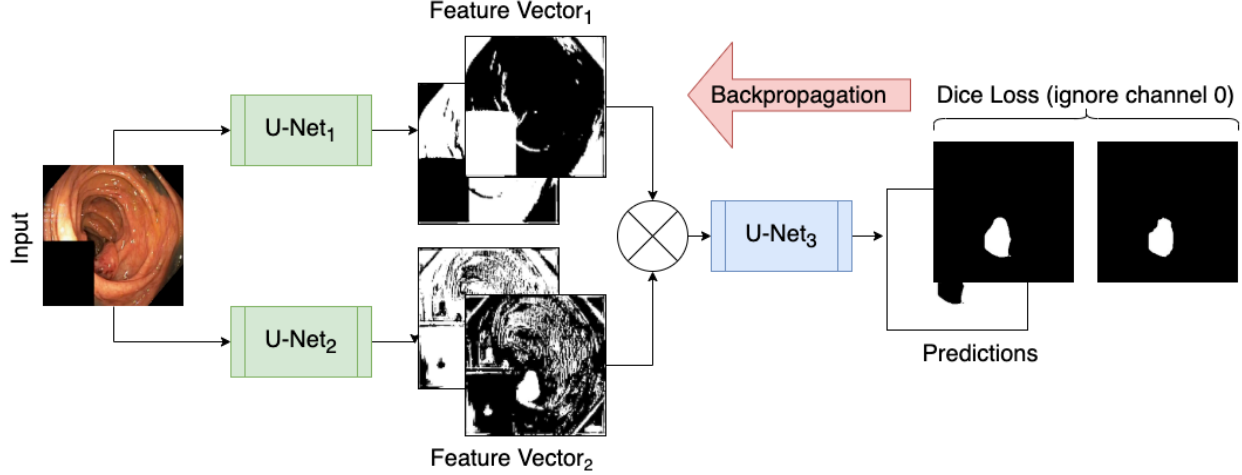


Fig. 3. TriUNet Architecture [24]

## 2) Precision

Precision is used to measure the positive patterns that are correctly predicted from the total predicted patterns in a positive class [26].

$$Precision(P) = \frac{tp}{tp + fp} \quad (6)$$

## 3) Recall

Recall is used to measure the fraction of positive patterns that are correctly classified [26].

$$Recall(r) = \frac{tp}{tp + tn} \quad (7)$$

## 4) IoU

The IoU (Intersection over Union) score is a standard performance measure for the object category segmentation problem. Given a set of images, the IoU measure gives the similarity between the predicted region and the ground-truth region for an object present in the set of images and is defined by following equation [27].

$$IOU = \frac{|A \cap B|}{|A \cup B|} \quad (8)$$

## V. RESULTS

### A. UNet

In this model, the `load_data` function loads the CVC612 dataset. Once the dataset is loaded, it is split into training, validation, and testing dataset using the function `train_test_split`.

The loss function is calculated using `binary_crossentropy`, and in the `build_model` function, the sigmoid is used for activation. The training was done for 20 epochs and the results are as follows: `loss: 0.1664`, `val_loss: 0.2206`, `accuracy: 0.9737`, `val_accuracy: 0.9435`, `recall: 0.8111`, `val_recall: 0.5573`, `precision: 0.9216`, `val_precision: 0.7460`, `IoU: 0.3326`, `val_IoU: 0.2248`. The evaluation of test results are as follows: `loss: 0.2483`, `accuracy: 0.9318`, `recall: 0.5968`, `precision: 0.6821`, `IoU: 0.2386`.

### B. TriUNet & DivergentNets

We have used TriUNet (combination of three UNets), UNet++, FPN, DeeplabV3 and DeeplabV3+ to implement DivergentNets. For training and validation, PolypGen dataset was used. Testing was done using Kvasir-Seg dataset. We implemented the experiments in Google Colab. Each model was implemented SE-ResNeXt-50-32x4D as the encoder [28], which was initialized with ImageNet [29] weights. The final prediction was produced by



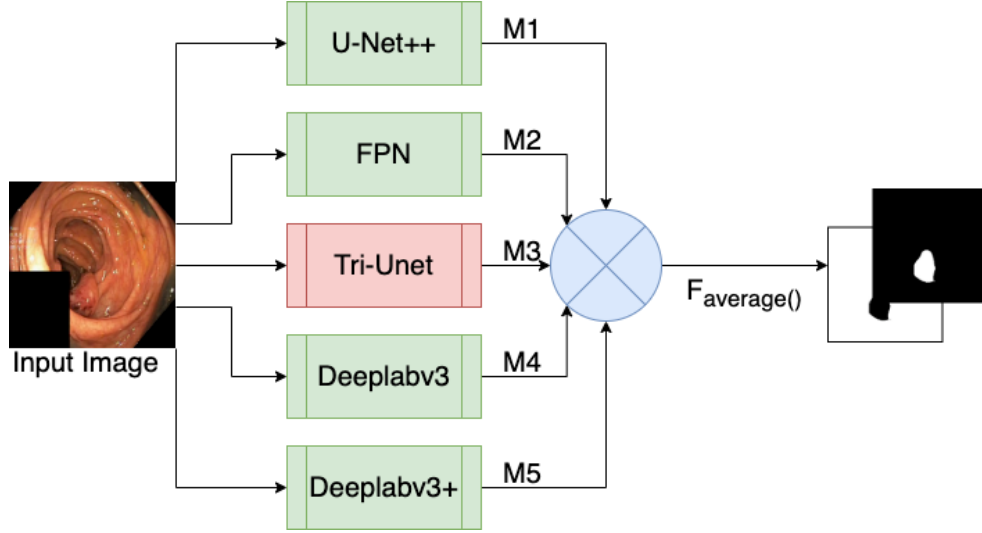


Fig. 4. DivergentNets Architecture [24]

passing the output through a two-dimensional softmax function. For training, all models started with a learning rate of 0.0001 and reduced to 0.00001 after 50 epochs. The model error was calculated using single-channel Dice for the polyp class, and the weights were optimized using Adam [30]. A series of different image augmentations were used to make the model more generalizable. These augmentations include horizontal flip, shift scale rotation, resizing, additive Gaussian noise, perspective shift, contrast limited adaptive histogram equalization (CLAHE), random brightness, random gamma, random sharpen, random blur, random motion blur, random contrast, and hue saturation. The augmentations were implemented using the Python library Albumentations [31]. No augmentations were applied to the validation and testing data. We tried color correction for testing on DivergentNets. The IoU before color correction was 0.7362 and after color correction, we got an IoU of 0.7378 as it is shown in Fig. 6. We also tried specular removal for testing and got an IoU of 0.8449. However, the IoU before specular removal was 0.9179 as it is shown in Fig. 7. After that, we tried different size of images and tested the IoU for both TriUnet and DivergentNets and the results are depicted in Fig. 8.

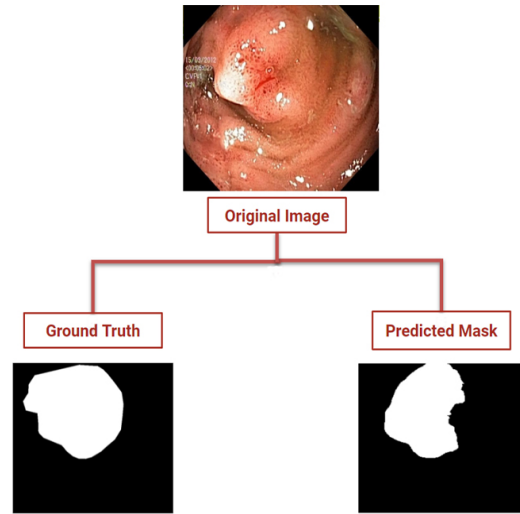


Fig. 5. DivergentNets Results

## VI. DISCUSSION

Our approach was to increase the performance of polyps segmentation by modifying or adding techniques. We have implemented the UNet model and the results are not that satisfactory. So, we have implemented two recent models, TriUnet and DivergentNets, which produced better results than UNet. We have improved the performance by adding some techniques like color correction Fig. 6. Also, we have tried to remove the specularity from



the images, but the IoU decreased as the polyps were also smoothed during the process Fig. 7. In addition, we have tested the models with different size of images and observed that as the resolution increases the IoU increases which means having more details can produce better prediction Fig. 8.

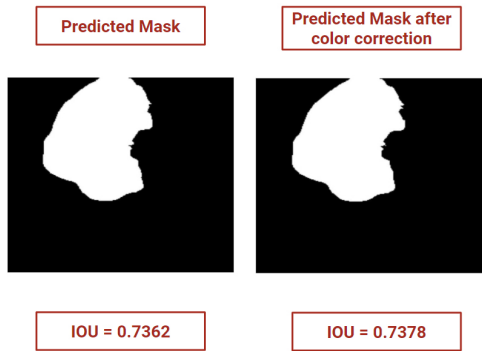


Fig. 6. Color Correction Results

	Images	Predicted Mask	IOU
Original Image			0.9179
Specularity Removed Image			0.8449

Fig. 7. Specular Removal Results

## VII. CONCLUSION

Segmentation of polyps for colorectal cancer is a challenging task because of the different shapes, colors, and intensities of the polyps. Moreover, another challenge would be the lack of access to a large public polyp dataset with annotations. In the current study, we presented our approaches to the polyp detection challenge. We used three models including UNet, TriUNet, DivergentNets for colorectal polyp segmentation. Comparison of all three methods

shows that a combination of a variety of segmentation models achieves a higher IoU score and detects polyps more precisely. In addition, we tried three other techniques, such as, color correction, image pyramid and specular removal to enhance the performance.

In future, we suggest changing the configuration of TriUNet and DivergentNets by integrating with other segmentation models for each node to enhance the performance. We also plan to try some other preprocessing techniques to achieve better results.

## ACKNOWLEDGMENT

The authors would like to thank Dr. Basu for his support to discuss the project and his comments.

## REFERENCES

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] M. B. Huck and J. L. Bohl, "Colonic polyps: Diagnosis and surveillance," *Clinics in colon and rectal surgery*, vol. 29, no. 04, pp. 296–305, 2016.
- [3] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-seg: A segmented polyp dataset," in *International Conference on Multimedia Modeling*. Springer, 2020, pp. 451–462.
- [4] A. Mishra, Y. Aloimonos, and C. Fermuller, "Active segmentation for robotics," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 3133–3139.
- [5] P. Garg, A. S. Chakravarthy, M. Mandal, P. Narang, V. Chamola, and M. Guizani, "Isdnet: Ai-enabled instance segmentation of aerial scenes for smart cities," *ACM Transactions on Internet Technology (TOIT)*, vol. 21, no. 3, pp. 1–18, 2021.



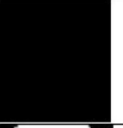













Image Size	Original	Ground Truth	Prediction for TriUnets	Prediction for DivergentNets	IOU (TriUnet)	IOU(DivergentNets)
32*32					3.4481e - 09	3.4483e - 09
64*64					0.5093	0.6061
128*128					0.7182	0.7471
256*256					0.7865	0.8188

Fig. 8. Comparing the Results of TriUNet and DivergentNets Models on One Image with Different Sizes

- [6] A. Sagar and R. Soundrapandiyan, "Semantic segmentation with multi scale spatial attention for self driving cars," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2650–2656.
- [7] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annual review of biomedical engineering*, vol. 2, no. 1, pp. 315–337, 2000.
- [8] S. Hwang, J. Oh, W. Tavanapong, J. Wong, and P. C. de Groen, "Polyp detection in colonoscopy video using elliptical shape feature," in *2007 IEEE International Conference on Image Processing*, vol. 2, Sep. 2007, pp. II – 465–II – 468.
- [9] S. Gross, M. Kennel, T. Stehle, J. Wulff, J. Tischendorf, C. Trautwein, and T. Aach, "Polyp segmentation in nbi colonoscopy," 01 2009, pp. 252–256.
- [10] D. Jha, P. Smedsrud, M. Riegler, D. Johansen, T. de Lange, P. Halvorsen, H. Johansen, and Simulamet, "Resunet++: An advanced architecture for medical image segmentation," 12 2019.
- [11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [12] Z. Zhang and Q. Liu, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. PP, 11 2017.
- [13] A. Galdran, G. Carneiro, and M. A. G. Ballester, "Double encoder-decoder networks for gastrointestinal polyp segmentation," in *International Conference on Pattern Recognition*. Springer, 2021, pp. 293–307.
- [14] H. M. Afify, K. K. Mohammed, and A. E. Hassanien, "An improved framework for polyp image segmentation based on segnet architecture," *International Journal of Imaging Systems and Technology*, 2021.
- [15] Y. Guo, J. Bernal, and B. Matuszewski, "Polyp segmentation with fully convolutional deep neural networks—extended evaluation study," *Journal of Imaging*, vol. 6, p. 69, 07 2020.
- [16] W. Cai, S. Chen, and D. Zhang, "Fast

- and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation,” *Pattern Recognition*, vol. 40, no. 3, pp. 825–838, 2007.
- [17] Z. Zhang, Q. Liu, and Y. Wang, “Road extraction by deep residual u-net,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.
- [18] D. Jha, S. Ali, N. K. Tomar, H. Johansen, D. Johansen, J. Rittscher, M. Riegler, and P. Halvorsen, “Real-time polyp detection, localization and segmentation in colonoscopy using deep learning,” *IEEE Access*, vol. PP, pp. 1–1, 03 2021.
- [19] E. Naess, V. Thambawita, S. A. Hicks, M. A. Riegler, and P. Halvorsen, “Pyramidal segmentation of medical images using adversarial training,” in *Proceedings of the 2021 Workshop on Intelligent Cross-Data Analysis and Retrieval*, 2021, pp. 33–38.
- [20] M. Akbari, M. Mohrekesh, E. Nasr-Esfahani, S. R. Soroushmehr, N. Karimi, S. Samavi, and K. Najarian, “Polyp segmentation in colonoscopy images using fully convolutional network,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 69–72.
- [21] 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 Medical Imaging and Graphics*, vol. 43, pp. 99–111, 2015.
- [22] S. Ali, D. Jha, N. Ghatwary, S. Realdon, R. Cannizzaro, O. E. Salem, D. Lamarque, C. Daul, K. V. Anonsen, M. A. Riegler *et al.*, “Polypgen: A multi-center polyp detection and segmentation dataset for generalisability assessment,” *arXiv preprint arXiv:2106.04463*, 2021.
- [23] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” 2014.
- [24] V. Thambawita, S. A. Hicks, P. Halvorsen, and M. A. Riegler, “Divergentnets: Medical image segmentation by network ensemble,” *arXiv preprint arXiv:2107.00283*, 2021.
- [25] A. Sanchez-Gonzalez, B. Garcia-Zapirain, D. Sierra-Sosa, and A. Elmaghraby, “Automated colon polyp segmentation via contour region analysis,” *Computers in biology and medicine*, vol. 100, pp. 152–164, 2018.
- [26] M. Hossin and M. N. Sulaiman, “A review on evaluation metrics for data classification evaluations,” *International journal of data mining & knowledge management process*, vol. 5, no. 2, p. 1, 2015.
- [27] M. A. Rahman and Y. Wang, “Optimizing intersection-over-union in deep neural networks for image segmentation,” in *International symposium on visual computing*. Springer, 2016, pp. 234–244.
- [28] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
- [29] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, “Imagenet large scale visual recognition challenge,” *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [30] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [31] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, “Albumentations: fast and flexible image augmentations,” *Information*, vol. 11, no. 2, p. 125, 2020.