Medico 2021: Medical Image Augmentation and Segmentation using Combination of Segmentation Neural Networks

Zeshan Khan, Mubasher Khan, Mubashir Yasin, Muhammad Hassan, Muhammad Atif Tahir
{zeshan.khan,p180111,p180143,p176157,atif.tahir}@nu.edu.pk
FAST School of Computing, National University of Computer and Emerging Sciences, Pakistan

ABSTRACT

Polyp identification is a critical task for the pre-detection of colon cancer. Proper removal of a polyp requires accurate estimation of the size and shape of the polyp. The identification of the size and shape of the polyp can be done using polyps segmentation. This research investigated various polyps segmentation approaches evaluated on the benchmark dataset of Kvasir. The best results by the use of UNet++ architecture on the augmented data resulted in an accuracy of 0.92 with a Dice coefficient of 0.53.

1 INTRODUCTION

Capsule endoscopy has been used for endoscopic abnormalities diagnostics for more than 10 years. Endoscopic images provide diagnosis capability for the detection of several abnormalities including various types of cancers in the Gastrointestinal Tract (GI-Tract), ulcer, and polyps detection. The analysis of such video frames takes a lot of time of medical experts, which can be reduced by the use of computer-aided diagnostics. with the increase in processing powers of computational machines, deep learning (DL) based automated diagnostics can result in good accuracy and efficiency.

2 RELATED WORK

The GI-Tract disease detection is an active area of research with the benefits of computer-aided diagnostics of various endoscopic diseases. There are various works on the segmentation of the polyps using neural network approaches.

Jha et al. [9] investigated the semantic segmentation of polyps in the GI-Tract. An auto-encoder-based architecture of ResUNet is used in the research for the segmentation of polyp. A modified version of the ResUNet with the name ResUNet++ was proposed [7].

Trinh et al. [13] used an auto-encoder-based approach with the replacement of RelU with the leaky ReLU. The Leaky-ReLU is a ReLU with some dead neurons enhances the results by ignoring some of the neurons in the computation of ReLU. The network for the encoder and decoder used by the authors is based on resnet50 trained on imagenet [5]. The approach resulted in 0.95 accuracies when tested on the MediaEval 2020 challenge dataset [8].

Brandao et al. [4] converted Convolution Neural Network (CNN) to Fully Connected Network (FCN) in their architecture for the segmentation of the polyps. The basic idea of the detection was deconvolution of the images before providing to the neural network for the detection of pixels if these are polyps or not.

3 APPROACH

The methodology of the research is data augmentation and segmentation. The data augmentation is done by applying several noises and reshaping methodologies on the images. The sequence of the image operations for the noise and reshape are crop, adding noise, horizontal and vertical flipping, mirroring, scaling, brightness change, contrast, and sharpness. The parameters for the various augmentation operations were random in a range such that the output image should be of size 224×224 with three channels. These operations resulted in augmented images. The augmentation is applied to generate 1300 images from the training set and the same operation is applied on the ground truths as well for the augmentation.

The segmentation of the image is done using various neural networks and clustering techniques. The methodologies were evaluated on the evaluation data, which was taken as 20% from training data. The methodologies of the neural network approaches worked better on the evaluation data than clustering techniques.

The neural network approaches for the segmentation of the images were based on auto-encoder architectures inspired by U-Net [12], U-Net++ [15], ResUNet++ [9] and SegNet [2]. The various auto-encoder approaches were evaluated on the validation dataset and the approaches of UNet++ architecture gave good accuracy and the Dice score. Based on the validation data, the best approaches used for the segmentation of the test data were a 10-layered UNet++ based auto-encoder and a ResUNet++ auto-encoder.

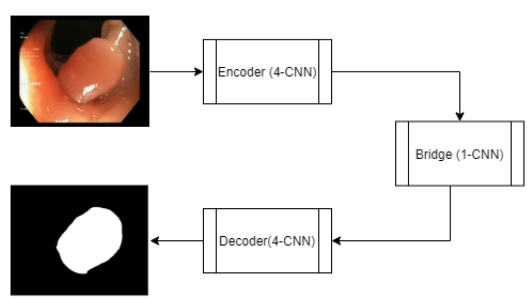

Figure 1: Architecture of Unet++.

The UNet++ auto-encoder has 4 CNN layers in the encoder and then a CNN layer as a Bridge then another set 4 of CNN with the same filters as in the encoder but in reverse. The auto-encoder output is again passed through a CNN layer to compute the masks

This research work was funded by Higher Education Commission (HEC) Pakistan under NRPU Project 10225/2017.

MediaEval'21, December 13-15 2021, Online

of the provided image. The architecture of the Unet++ is shown in Figure 1. The auto-encoder is trained for the 100 epochs with a batch size of 25 based on the availability of the resources. These 100 epochs training with loss function of binary cross-entropy and learning rate of 0.0001 resulted in the best validation Dice and accuracy.

The auto-encoder build on the inspiration of the ResUNet++ [9] consists of encoder and decoder using residual blocks in the network. The encoder used one stem and three residual blocks of three convolutions in each of the residual and a stem block with some pooling and the fully connected layers. There is a similar decoder block with the same number of convolution and pooling layers with the addition of the attention layers in between them. The decoder's convolution layers are in a similar pattern as in the encoder with the reverse order. The model is trained for the 200 epochs with the learning rate of 0.0001 with a batch size of 8 images. Similar to the ResNet++, the loss function of the binary cross-entropy was used with the Adam optimizer.

The overall methodology of the training can be expressed as the data augmentation and the segmentation by the Figure 2.

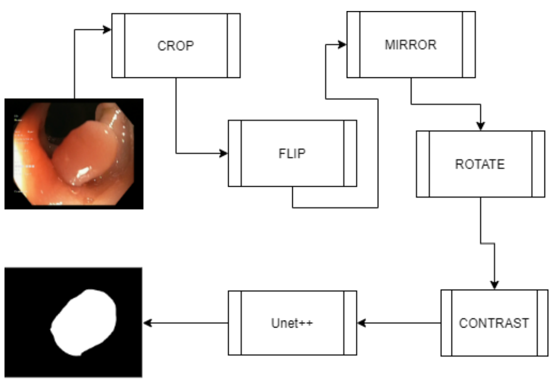

Figure 2: System Architecture.

4 DATASET

The dataset for the research is provided by the Simula Lab as Hyper Kvasir dataset [3] for the training and the test data as MediaEval 2021 [6]. This data consists of 1360 RGB images for the segmentation of the polyps with the ground truth masks for these images. The images were of various dimensions from 352×449 to 1072×1024 with RGB channels. The test data consists of 200 unlabeled RGB images of sizes varying from 576×576 to 1072×1072.

5 RESULTS AND ANALYSIS

The various approaches for the polyps segmentation were done on the training dataset with the 30% validation data from training. The various image segmentation networks and the pixel clustering approaches were investigated with the 70% training and 30% validation dataset for the evaluation measure of accuracy and Dice coefficient. The task rules were the submission of the 5 best runs. So, the top 5 methodologies based on validation results were selected for the submission. The official results from organizers show the evaluations as in Table 1. The best results were received by the application of the Unet++ with Adam optimizer on augmented images.

Table 1: Test Data Results

Model	Dice	Accuracy	Jaccard Index	P	R
UNet++ With Augmentation	**0.53**	**0.92**	**0.42**	**0.66**	**0.53**
UNet++ Without Augmentation	0.34	0.88	0.25	0.49	0.36
ResNet++ With Augmentation	0.5	0.91	0.40	0.64	0.50
ResNet++ Without Augmentation	0.34	0.88	0.25	0.45	0.37
Averaged All With Augmentation	0.44	0.91	0.34	0.65	0.40

The P in Table 1 is used for the precision and R for the recall. Table 1 shows a comparison of the test results for the top 5 approaches applied for the polyps segmentation. There are three evaluation measures used in the comparison. The Dice coefficient, pixel accuracy, and the Jaccard Index. The Dice coefficient is the two times ratio of the common area of both images over the total area of both the images. The incorrect segmentation reduces the common area a lot and affects the Dice coefficient a lot. So, the Dice coefficient of the various approaches is 0.34 to 0.53 only. The pixel accuracy is the proportion of the pixels classified correctly over the total number of pixels. The polyps segmentation task has less than 20% of the image as polyps and the rest 80% image is not the polyps. So, most of the approaches where the polyps were present but not detected correctly caused a lesser impact on accuracy and the accuracy remained higher than 0.88. The Jaccard is a measure of the intersection over the union. The incorrectly segmented images can miss the huge portion of the intersection, polyps, that causes a low Jaccard Index.

6 CONCLUSION AND FUTURE WORK

The research is conducted using augmentation of the polyps images and segmentation using CNN-based auto-encoder architecture. The results of the approach show a good segmentation accuracy for the validation results. The results of the segmentation for some of the endoscopic images are not correct and the segmented region for polyps in those images is of zero pixels. This problem of no detection can be resolved by the application of detection before segmentation. In the future, we will investigate the detection as a pre-step for the segmentation of the polyps images based on Kvasir polyps detection datasets [10, 11]. The second problem in the results is low segmentation accuracy for the images with light reflections on the polyp [1, 14] which will be investigated for reflection removal methodologies to improve the segmentation accuracy.

REFERENCES

[1] Mojtaba Akbari, Majid Mohrekesh, Kayvan Najariani, Nader Karimi, Shadrokh Samavi, and SM Reza Soroushmehr. 2018. Adaptive specular reflection detection and inpainting in colonoscopy video frames. In *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 3134–3138.

[2] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 39, 12 (2017), 2481–2495.

[3] Hanna Borgli, Vajira Thambawita, Pia H Smedsrud, Steven Hicks, Debesh Jha, Sigrun L Eskeland, Kristin Ranheim Randel, Konstantin Pogorelov, Mathias Lux, Duc Tien Dang Nguyen, et al. 2020. HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific data* 7, 1 (2020), 1–14.

[4] Patrick Brandao, Odysseas Zisimopoulos, et al. 2018. Towards a computed-aided diagnosis system in colonoscopy: automatic polyp segmentation using convolution neural networks. *Journal of Medical Robotics Research* 3, 02 (2018), 1840002.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[6] Steven Hicks, Debesh Jha, Vajira Thambawita, Hugo Hammer, Thomas de Lange, Sravanthi Parasa, Michael Riegler, and Pål Halvorsen. 2021. Medico Multimedia Task at MediaEval 2021: Transparency in Medical Image Segmentation. In *Proceedings of MediaEval 2021 CEUR Workshop*.

[7] Debesh Jha, Sharib Ali, Nikhil Kumar Tomar, Håvard D Johansen, Dag Johansen, Jens Rittscher, Michael A Riegler, and Pål Halvorsen. 2021. Real-time polyp detection, localization and segmentation in colonoscopy using deep learning. *Ieee Access* 9 (2021), 40496–40510.

[8] Debesh Jha, Steven A Hicks, Krister Emanuelsen, Håvard Johansen, Dag Johansen, Thomas de Lange, Michael A Riegler, and Pål Halvorsen. 2020. Medico multimedia task at mediaeval 2020: Automatic polyp segmentation. *arXiv preprint arXiv:2012.15244* (2020).

[9] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Dag Johansen, Thomas De Lange, Pål Halvorsen, and Håvard D Johansen. 2019. Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)*. IEEE, 225–2255.

[10] Konstantin Pogorelov, Michael Riegler, Pål Halvorsen, Steven Hicks, Kristin Ranheim Randel, Duc Tien Dang Nguyen, Mathias Lux, Olga Ostroukhova, and Thomas de Lange. 2018. Medico multimedia task at mediaeval 2018. In *CEUR Workshop Proceedings*, Vol. 2283. Technical University of Aachen, 1–4.

[11] Michael Riegler, Konstantin Pogorelov, Pål Halvorsen, Carsten Griwodz, Thomas Lange, Kristin Randel, Sigrun Eskeland, Duc-Tien Dang-Nguyen, Mathias Lux, and Concetto Spampinato. 2017. Multimedia for medicine: the medico task at mediaeval 2017. (2017).

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.

[13] Quoc-Huy Trinh, Minh-Van Nguyen, Thiet-Gia Huynh, and Minh-Triet Tran. 2020. HCMUS-Juniors 2020 at Medico Task in MediaEval 2020: Refined Deep Neural Network and U-Net for Polyps Segmentation. (2020).

[14] Rui Yao, Yilun Wu, Wei Yang, Xiaolin Lin, Shidan Chen, and Su Zhang. 2010. Specular reflection detection on gastroscopic images. In *2010 4th International Conference on Bioinformatics and Biomedical Engineering*. IEEE, 1–4.

[15] Zhou Z., Rahman Siddiquee M.M., Tajbakhsh N., and Liang J. 2018. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *IEEE transactions on pattern analysis and machine intelligence* 11045 (2018). https://doi.org/10.1007/978-3-030-00889-5_1