International Journal of Engineering Works
ISSN-p: 2521-2419
ISSN-e: 2409-2770
Vol. 8, Issue 04, PP. 132-137, April 2021
https://www.ijew.io/
https://doi.org/10.34259/ijew.21.804132137

A Modified Memory-Efficient U-Net for Segmentation of Polyps

Asif Ahmad[1], Noor Badshah[2], Mahmood Ul Hassan[3]
[1,2,3] Department of Basic Sciences, University of Engineering and Technology Peshawar, Pakistan
asifahmad7007@gmail.com[1], noor2knoor@gmail.com[2], mahmoodulhassan300@gmail.com[3]
Received: 25 February, Revised: 28 March, Accepted: 14 April

***Abstract*—**Colorectal cancer, caused by an unusual growth of tissues in a body called polyp, is the third most prevailing cancer worldwide and remained the second most cause of deaths by cancer in 2020. Early stage detection of the cancer can prevent the deaths. Computer Aided Diagnosis (CAD) system could be a major breakthrough for early detection of the cancer. The system uses image processing techniques. Among the image processing techniques segmentation has a great value. The diagnostic process results are highly dependent on the accuracy of performed segmentation. Nowadays, many supervised and unsupervised techniques are used for the task of segmentation. Deep neural networks have outperformed other state-of-the-art approaches for the task. In this paper, we present an end-to-end deep neural network for segmentation of polyps in images. The network is modified version of the U-Net architecture. The network being much more memory efficient than the U-Net architecture, inferences segmentation of the images more accurate than the U-Net. We reduce number of layers of the U-Net architecture both in the en- coding and decoding path, and introduce residual blocks and batch normalization in the encoding path to prevent learning of redundant features, to avoid over-fitting and to accelerate the training process, and in the decoding path to avoid gradient vanishing issue in long dependence of the neural network during training we use bi-directional long short term memory network with batch normalization. We train and validate the network on Kvasir dataset for the task. The network accurately segments the polyp part in the images with 92.46% test accuracy.

Keywords: Deep Learning; Deep Neural Network; Image Segmentation; U-Net.

I. INTRODUCTION

Around 18.1 million new cancer cases and 10 million deaths by cancer occurred in 2020, according to GlOBOCAN 2020. Among the cancer cases female breast cancer is the most occurring cancer followed by lungs and colorectal cancers. In terms of deaths lung cancer remained the leading cancer with around 1.8 million deaths followed by 0.9 million deaths by colorectal cancer. Colorectal cancer is caused by an unusual growth of tissues known as polyp. Polyps are like moles, which are further developed into cancer. Early stage detection of the polyps can prevent the deaths by huge margin. Computer Aided Diagnosis (CAD) system could be a major breakthrough for early detection of the cancer [1]. CAD uses images processing techniques for analysis of medical images. Segmentation has a great value in the images processing. It divides the images into affected and unaffected region. The diagnostic process results are highly dependent on the accuracy of performed segmentation. Nowadays, many supervised and unsupervised techniques are used for the task of segmentation. Supervised techniques use active contours [2,3], fuzzy sets [4,5] and machine learning algorithms like k-mean clustering [6], morphological oper- ations [7], etc. While, unsupervised techniques come under deep learning. Deep learning uses convolutional neural networks (ConvNet/CNNs) [8–10]. Over the couple of years CNNs have outperformed other approaches in this field.

Since the emergence of Artificial Intelligence (AI) in 1950s, computer scientist have been trying to build a computer that can mimic human behavior. In the following decades the field saw much advancement. But they were limited due to unsophisticated computer and unavailability of large data. CNNs are special type of Artificial Neural Network (ANN), they are used to mimics the human vision system. CNNs were first designed and developed by Yann LeCun in 1980s [11]. The network was named LeNet after LeCun and it was trained for recognition of handwritten digits in banks and postal services. A breakthrough in the field was achieved in 2012 by AI system, known as AlexNet, developed by Alex Krizhevsky [12]. The system won the 2012 ImageNet computer vision contest with 85% accuracy.

In this paper, we propose a modified memory-efficient U-Net architecture for segmentation of medical images containing

polyps, as shown in figure 5. Our network is modification of the U-Net architecture 1. To make the network memory efficient we have made changes in the encoding and decoding path of the network. The paper is further organized as: section 2 touches related work, section 3 describes the proposed work in detail, in section 4 we present results of our network and comparison with the U-Net and the paper ends with a conclusion note in section 5.

II. Related Work

After the success of CNNs in computer vision application, they are being used in medical field for image processing purpose. In [13], the authors have looked into the possibility of direct use of CNNs for the segmentation of tumor tissues in brain. They have used inhomogeneity correction in each channel as a pre-process for the BraTS 2013 data. A total of 6 layers of convolutions are being used with max-pooling, softmax, and fully connected layers.

Small kernel of size 3x3 have been used in [14] for automatic CNN based segmentation of brain tumor. The method consists of three steps: pre-processing, classification by CNN, and post processing. Pre-processing stage contains bias-field correction and the architecture consists of convolution layers, max pooling and fully connected layers. While, in the post processing stage they have used volumetric constrains for the removal of erroneously segmented parts, which are smaller than the pre-defined threshold.

A major breakthrough achieved in the field in 2015 by U-Net architecture, figure 1, state-of- the-art architecture for segmentation of medical images. The U-Net, presented by Ronneberger et al. in 2015, is a state-of-the-art fast trained network based on fully convolutional network [8]. The work is known as U-Net due to its architecture shape, a symmetric 'U' shape. The network consists of two paths: an encoder/contraction path and a decoder/expansion path, which are on the left and right side of the model, respectively.

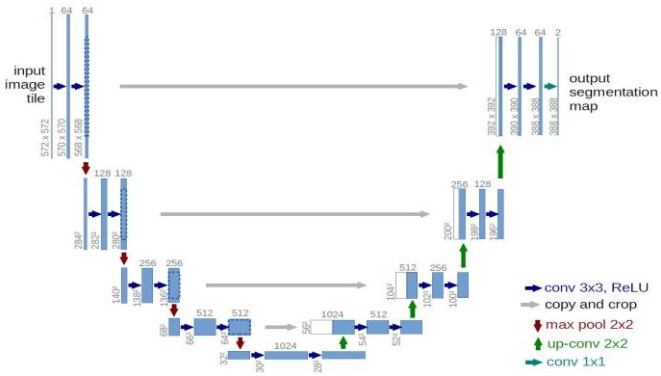

Figure 1: The U-Net architecture

The contraction path captures context and a symmetric expansion path enables precise localization. The network consists of total 23 layers of convolutions. Each second layer of 3x3 convolution, in the contraction path, is followed by a Rectified linear unit (ReLU) and a 2x2 max pooling operator with stride 2 for down sampling. And at each down-sampling the network makes the number of features doubled. While, in the expansion path up-sampling is followed by a 2x2 convolution, which at the same time halves the number channels, and consecutive two 3x3 convolution, where each layer is followed by a ReLU. At last, a final layer of 1x1 convolution is used to get the desired number of classes from each 64-component feature.

Due to the un-padding convolution layers in the contraction path the output resultant image of the network lose its boundary pixel, as example is given below in Figure 2. For the output in the yellow region the input object should be in the blue region, as the result of use of un-padding convolution. The missed part is gained then by using mirror extracting.

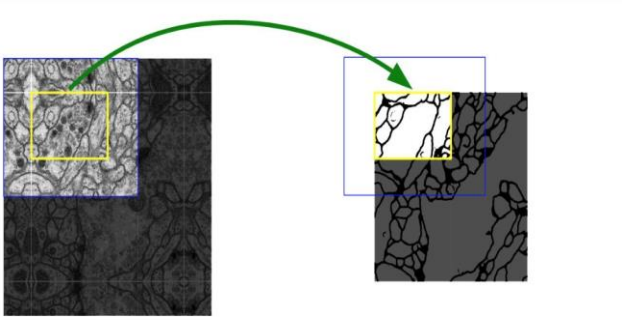

Figure 2: Seamless segmentation of arbitrarily large images by an overlap tile strategy by U-Net.

The authors of U-Net have applied the network in 2015 for segmentation task on datasets pro- vided by ISBI cell tracking challenge, the challenge began in 2012 and still available for challenge. In the first, they applied the network on dataset, PhC-U373, which contained Glioblastoma-astrocytoma U373 cells. It obtained 92 % of average value of intersection over union (IOU), and on the second dataset, DICHeLa, they achieved an average 77.5 % of IOU value. In both cases they got first position.

In [15], the authors have presented 'Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks', see Figure 3. Unlike the U-net, they have used zero-padding, which keeps the output dimension for all the convolutional layers of both down-sampling and up-sampling path. Besides the shape of U-Net, it also comprises of 23 layers and the function of the convolution layers, ReLU and max pooling are in the same order. Along the zero-padding, they have also used 3x3 convolution for up-sampling, unlike the U-Net.

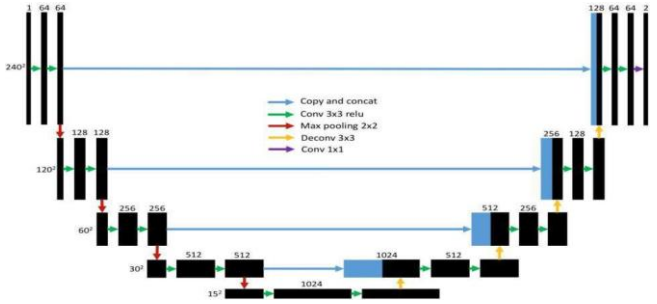

Figure3:The automatic brain tumor segmentation architecture based on the U-Net.

A 3D improved U-Net based network is presented by authors of [16]. The proposed network comprises of 3x3 convolutions, group normalization, ReLU, dropout and max-pooling layers, which are used in the contraction path. And in the expansion path 2x2 convolution layers for up-sampling, 3x3 convolution, group normalization, and ReLU are used. At last, 1x1 convolution is followed by a softmax layer. In the presented work the authors have generated heatmaps of different types of lesions by utilizing ground-truth of brain tumor from group of patients. Then, volume of interest (VOI) is created by these heatmaps, which contains advance information of brain tumor lesions. The multimodal MRIs are then integrated with VOI map and is used as input for the network, as shown below:

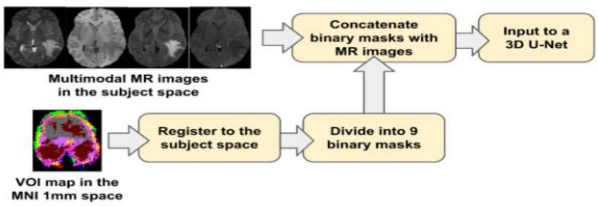

Figure 4: Pipeline of 3D U-Net

In this paper, we propose a modified memory-efficient U-Net architecture for segmentation of medical images containing polyps, as shown in figure 5. Our network is modification of the U-Net architecture 1. To make the network memory efficient we have made changes in the encoding and decoding path of the network. The next section 3 presents the proposed network in detail.

III. PROPOSED NETWORK

In this paper, we propose a moified memory-efficient U-Net architecture for segmentation of medical images containing polyps, as shown in figure 5. Our network is modification of the U- Net architecture. To make the network memory efficient we have made changes in the encoding and decoding path of the network. We introduce residual blocks [17] and batch normalization [18] to prevent learning of redundant features, over-fitting, acceleration of the training process, and to reduce number of parameters which ultimately leads to lower computational cost; and in the decoding path we take benefit of bi-directional long short term memory network [19] to avoid gradient vanishing issue in long dependence of the neural network during training.

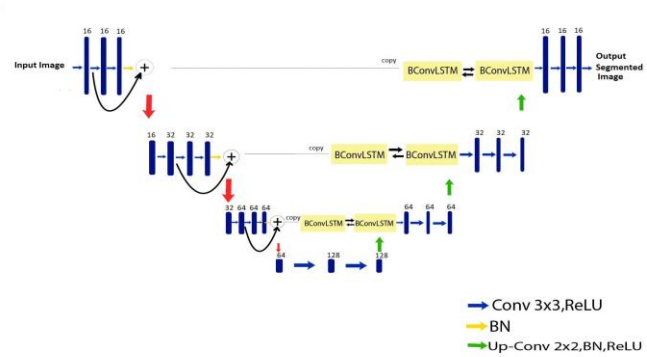

Figure 5: Our network. Comparing with the U-Net one can see the less number of layers and other modifications in the network.

The working mechanism of a CNN can be represented in equations as: If x represents an input, f represents activation function in a layer, W^n represents a kernel/filter, used to convolve over an image in a convolution operation, in an nth layer, b^n represents bias term added in the nth layer, z^n represents an output of a convolution operation with bias term in an nth layer, and a^n represents an output of the f in an nth layer, then equations for n number of layers can be:

$$z^1 = W^1 x + b^1$$

$$a^1 = f(z^1)$$

$$a^1 = f(W^1 x + b^1)$$

$$z^2 = W^2 \, f(W^1 x + b^1) + b^2$$

$$a^2 = f(W^2 \, f(W^1 x + b^1) + b^2)$$

Similarly, third layer activation function output can be represented by:

$$a^3 = f(W^3 f(W^2 f(W^1 x + b^1) + b^2) + b^3)$$

Likewise, nth layer activation function output can be represented as:

$$a^n = f(W^n f(W^{n-1} \ldots f(W^1 x + b^1) + \ldots + b^{n-1}) + b^n)$$

Convolutional neural network (CNN) based neural architectures are build up of convolution operation, pooling operation, activation function, etc.

In the encoding path of our network, we use residual blocks mechanism followed by batch normalization layer. Residual blocks can be represented as: if X represents an input and F represents convolution operation and activation function in a single block then

$$X = X + F(X)$$

represents a single application of residual blocks. That is, an input to some layers of convolution, activation function or max-pooling is again added to the output of the layers. In our network, we use residual blocks to two layers of convolution followed by activation function as shown in the figure 5 above. The output of the residual block again passes through batch normalization layer, the equation for batch normalization is given as:

$$BN = \gamma_c \left[\frac{I_{n,c,h,w} - \mu_c}{(\sigma_c^2 - \varepsilon)^{1/2}} \right] + \beta_c$$

Where, $I_{n,c,h,w}$ represents n-number of images provided to a neural network at a time with c channels, h heights and w widths. μ_c and σ_c^2 are channel wise global mean and variance of the images, respectively. β_c and γ_c are learnable mean and standard deviation, respectively, while ε is kept constant as 0.00001. Batch normalization layer controls variation in distribution by calculating mean and standard deviation values of the data set as a whole by adjusting the mean to 0 and variance to 1.

While in the decoding path, we use bi-directional long short term memory network (BConvLSTM) preceded by again batch normalization layer and followed by three convolution layers, as can be seen in right side of the network given in figure 5. LSTMs are modified version of recurrent neural networks (RNNs) [20], which have been developed to overcome the gradient vanishing issue in long dependence of neural networks in training.

IV. Training And Results Comparison

We take advantage of free graphic processing unit (GPU) provided by Google Colab for training and evaluation of the network for the task. We use Kvasir-SEG dataset for training and evaluation. The dataset contains 1000 images and respective 1000 masks, where each image contains polyp(s). The dataset comprises gastrointestinal (GI) tract images and was released for 2020 MediaEval Medico-polyp segmentation. We do not use any pre-processing techniques for the training process. We just use four type of data augmentation, RandomRotate90, GridDistortion, HorizontalFlip and VerticalFlip, to increase size of the dataset from 1000 to 5000 images and respective masks. We then distribute the dataset into 4000 images for training, 1000 for validation and 1000 for testing. All images in the dataset are then resized into 256x256 images to accommodate the training process in the GPU. The model is being developed in Keras with TensorFlow backened framework.

We start the training process with batch size 8, learning rate starting from 10^{-3}, we let the learning rate to reduce to 10^{-5} in case of not improvement in validation loss with factor 0.1 and patience 7, and 100 epochs. The network trained for 31 epochs instead of 100 epochs and the training process stopped there because of not improvement of results. Visualization of the training process which shows the model accuracy and loss given in the figure below 6.

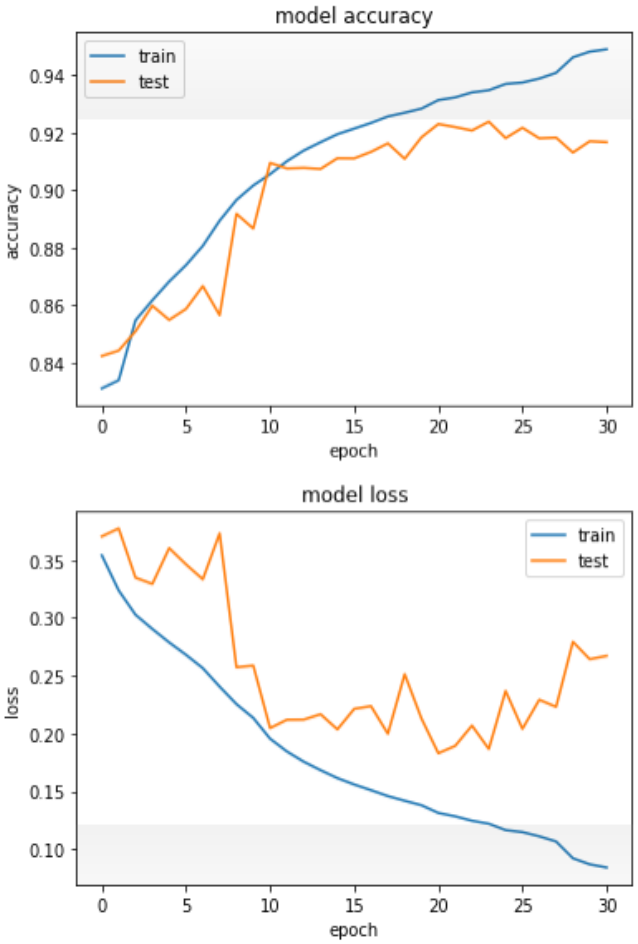

Figure 6: Training and validation accuracy and loss of our network.

Over-fitting of the training of the model can be seen in the figures above. This is due to the random selection of data for training and validation, and also small size of the dataset. During the training we achieve 94.90% and 08.43% training accuracy and loss, respectively, and 92.38% and 18.30% validation accuracy and loss, respectively. While, we achieve 92.46% and 17.71% testing accuracy and loss, respectively. Some of the results of testing of the model on the training images are given below in figure 7.

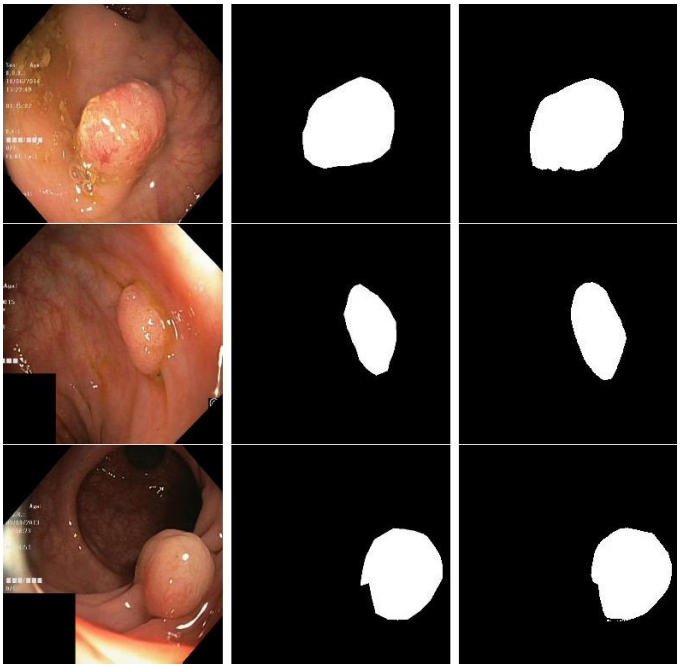

Figure 7: Segmentation results of the proposed network. Left: Image, Middle: Image mask, Right: Our network prediction.

To mitigate the over-fitting, we add 5% Gaussian noise to the input images and add dropout layers between the blocks of encoding path of the network. We also compare our results quantitatively with the original U-Net in term of accuracy, parameters and GPU time in the table 1 below.

Table 1: Quantitatively comparison of our network results with the U-Net.

Models	**Accuracy (%)**	**Parameters**	**Time Per Epoch(sec)**
Our Network	92.46	618,549	181
U-Net	91.83	2,274,501	308

We kept the same parameters for training of the U-Net on the dataset. The table shows that our network has achieved higher accuracy with much less parameters and computational cost as compare to the U-Net architecture. Experimental results show that if we provide large size of dataset to our network we can achieve much more accurate results and which will ultimately lead to good fit of the network for the task.

CONCLUSION

We presented a deep memory efficient neural network for segmentation of polyps. The network segmented the polyps more accuractely than the U-Net architecture with less number of paramters and quite less computational cost. It would be not unfair to say that U-Net architecture has brought a revolution in the segmentation of medical images in the field of deep learning. The network has achieved outstanding results on different kind of biomedical images. The idea has been used for different biomedical applications with some minute modifications. There are other options too for the advancement in current architecture. We can make the network even more fast and can also train for big and new datasets like ImageNet and BraTS 2018, which would improve its performance adequately.

REFERENCES

[1] C. Dromain, B. Boyer, R. Ferre, S. Canale, S. Delaloge, and C. Balleyguier, "Computed-aided diagnosis (cad) in the detection of breast cancer," *European journal of radiology*, vol. 82, no. 3, pp. 417–423, 2013.

[2] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.

[3] L. Wang, L. He, A. Mishra, and C. Li, "Active contours driven by local gaussian distribution fitting energy," *Signal Processing*, vol. 89, no. 12, pp. 2435–2447, 2009.

[4] O. J. Tobias and R. Seara, "Image segmentation by histogram thresholding using fuzzy sets,"

[5] A. Ahmad, N. Badshah, and H. Ali, "A fuzzy variational model for segmentation of images having intensity inhomogeneity and slight texture," *Soft Computing*, pp. 1–16, 2020.

[6] C. W. Chen, J. Luo, and K. J. Parker, "Image segmentation via adaptive k-mean clus- tering and knowledge-based morphological operations with biomedical applications," *IEEE transactions on image processing*, vol. 7, no. 12, pp. 1673–1683, 1998.

[7] S. Z. Oo and A. S. Khaing, "Brain tumor detection and segmentation using watershed seg- mentation and morphological operation," *International Journal of Research in Engineering and Technology*, vol. 3, no. 03, pp. 367–374, 2014.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical im- age segmentation," in *International Conference on Medical image computing and computer- assisted intervention*, pp. 234–241, Springer, 2015.

[9] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional convlstm u-net with densley connected convolutions," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 0–0, 2019.

[10] L. Zhang, A. Liu, J. Xiao, and P. Taylor, “Dual encoder fusion u-net (defu-net) for cross- manufacturer chest x-ray segmentation,” arXiv preprint arXiv:2009.10608, 2020.

[11] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition,” Neural computation, vol. 1, no. 4, pp. 541–551, 1989.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolu- tional neural networks,” in Advances in neural information processing systems, pp. 1097– 1105, 2012.

[13] D. Zikic, Y. Ioannou, M. Brown, and A. Criminisi, “Segmentation of brain tumor tissues with convolutional neural networks,” Proceedings MICCAI-BRATS, pp. 36–39, 2014.

[14] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain tumor segmentation using convolutional neural networks in mri images,” IEEE Transactions on Medical Imaging, vol. 35, pp. 1240–1251, May 2016.

[15] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, “Automatic brain tumor detection and seg- mentation using u-net based fully convolutional networks,” in annual conference on medical image understanding and analysis, pp. 506–517, Springer, 2017.

[16] P.-Y. Kao, J. W. Chen, and B. Manjunath, “Improving 3d u-net for brain tumor segmenta- tion by utilizing lesion prior,” arXiv preprint arXiv:1907.00281, 2019.

[17] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770– 778, 2016.

[18] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by re- ducing internal covariate shift,” arXiv preprint arXiv:1502.03167, 2015.

[19] H. Song, W. Wang, S. Zhao, J. Shen, and K.-M. Lam, “Pyramid dilated deeper convlstm for video salient object detection,” in Proceedings of the European conference on computer vision (ECCV), pp. 715–731, 2018.

[20] M. I. Jordan, “Attractor dynamics and parallelism in a connectionist sequential machine,” in Artificial neural networks: concept learning, pp. 112–127, 1990.

How to cite this article:

Asif Ahmad, Noor Badshah, Mahmood Ul Hassan “A Modified Memory-Efficient U-Net for Segmentation of Polyps”, International Journal of Engineering Works, Vol. 8, Issue 04, PP. 132-137, April 2021, https://doi.org/10.34259/ijew.21.804132137.