

# Transfer of Knowledge: Fine-tuning for Polyp Segmentation with Attention

Rabindra Khadka<sup>1</sup>

<sup>1</sup>Simulamet, Norway

rabindra.khadka@ymail.com

## ABSTRACT

This paper describes how the transfer of prior knowledge can effectively take on segmentation tasks with the help of attention mechanisms. The UNet model pretrained on brain MRI dataset was fine-tuned with the polyp dataset. Attention mechanism was integrated to focus on relevant regions in the input images. The implemented architecture is evaluated on 200 validation images based on intersection over union and dice score between ground truth and predicted region. The model demonstrates a promising result with computational efficiency.

## 1 INTRODUCTION

Early detection of polyps is vital to reduce colorectal cancer (CRC) deaths as it is one of the most common types of cancer reported in the world [11]. Colonoscopy is performed in order to detect polyps in the gastrointestinal tract, which assists doctors to perform timely intervention before they become malign. Acknowledging the significance of having accurate segmentation techniques in a clinical setting, *MediaEval Challenge 2020* organized the automatic polyp segmentation challenge to develop systems that efficiently detect and segment colon polyps [3].

Machine learning algorithms like deep learning models for semantic image segmentation have recently shown promising results in medical setting [13, 14]. So, machine learning models can be used to aid doctors while performing endoscopy that can bring potential polyp segment into doctor's attention, which could have been missed or incorrectly passed. In this work, the knowledge transfer approach was adapted using a pre-trained UNet guided with attention mechanism [9]. The pre-trained network was coupled with attention and then fine-tuned to achieve faster convergence and a good validation score.

## 2 RELATED WORK

Traditionally, a number of manual techniques to extract polyp features such as color, shape, appearances have been used to train a classifier to identify polyps from its background [12]. With the advent of deep learning models, polyp segmentation problem has been approached by learning polyp and its mask. ResUNet++ [4] for polyp segmentation shows a promising result which is based on deep residual UNet (ResUNet) structure. The other work includes PraNet with complex architecture that takes into account the relationship between polyp area and its boundary [2] which used reverse attention mechanism to model the boundaries of the polyps. Their work showed improved results on various data-sets.

Attention mechanism has been commonly used in NLP domain. There are two types of trainable attention mechanism namely Hard attention [8] and soft attention [1]. Self-attention mechanism also have been proposed which eliminates the dependency on external gating information and have shown better performance results. Self-attention mechanism has been used with UNet for segmentation of medical images(pancreas,abdominal) that showed promising state of art results across different datasets [9].

## 3 APPROACH

### 3.1 Data

Data used for this challenge was Kvasir-SEG [5] dataset which consist of 1,000 polyp images and the respective ground truth. The data set was divided into training and validation set in the ratio of 80:20. Test data set consisted of 160 images without ground truth.

### 3.2 Data processing

Images were normalized in the range of  $[-1,1]$  and resized to  $256 \times 256$ . Augmentation techniques were also applied to the images randomly before feeding them into the model; namely horizontal flipping, vertical flipping, mirroring, rotation(-5 to 5), elastic transformation, channel shifting and solar flares.

### 3.3 Model Architecture

UNet is known for its efficient performance on image segmentation task[10]. It provides the architecture with multi stage cascaded convolutions neural networks which helps to extract the region of interest and make dense prediction. Despite UNet having a good representational power, it also redundantly uses compute resources as it repeatedly extracts low level features. Therefore, for overcoming this drawback of UNet, attention mechanism can be integrated with the UNet architecture. This has led to improvement of model's sensitivity to region of interest and also suppresses features response from irrelevant regions in the image. The soft additive attention have shown better performance than the multiplicative attention [7].

In this work, the notion of knowledge transfer has been the key motivation factor to choose a simple pre-trained model. The additive soft attention mechanism was integrated with the pre-trained UNet architecture. The chosen pre-trained model has been trained on brain MRI dataset [6]. The schema of the *attention UNet* architecture can be seen Figure 1. The key benefit of this attention UNet structure in compare to multi stage CNNs is that it does not require training of multiple models to deal with object localization and thus reduces number of model parameters. As seen in Figure 1, additive attention is applied to obtain the gating coefficient  $\alpha$ .

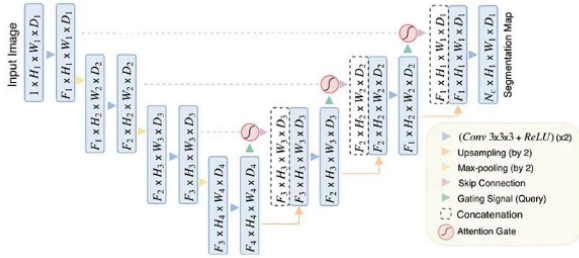


Figure 1: Schema of Attention with U Net segmentation model. Input image is down sampled by the factor of 2 in the encoding section of the model. Attention gates are connected with skip connection which filters the features. [9]

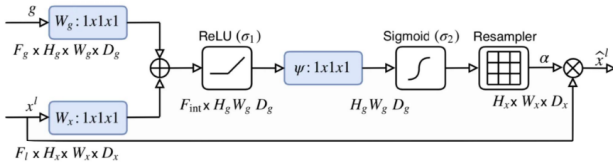


Figure 2: Schema of Attention block. Two input features are  $x_i$  and  $g$  are passed through conv. block and then scaled by attention factor  $\alpha$ . Resampler is a trilinear interpolation. The output of the block is concatenated with the upsampled feature maps at lower level of the model. [9]

### 3.4 Loss function

A compound loss was used during training which comprises of both dice loss ( $L_{Dice}$ ) and binary cross entropy loss ( $L_{BCE}$ ) as ( $Loss = L_{BCE} + L_{Dice} + \lambda ||\theta||_2^2$ ).  $L_2$  regularization was applied while optimizing the loss.

$$L_{Dice} = 1 - \frac{2 * \sum y_{true} * y_{pred}}{\sum y_{true}^2 + \sum y_{pred}^2 + \epsilon} \quad (1)$$

$$L_{BCE}(y, \hat{y}) = -(y \log P(\hat{y}) + (1 - y) \log(1 - \hat{y})) \quad (2)$$

where  $y$  and  $\hat{y}$  are the ground truth value and predicted value respectively.

### 3.5 Implementation details

The model was implemented using pytorch. It was initially trained on single GPU from google's colab. Gradient updates were performed with the batch size of 16. The initial learning rate of  $1e - 4$  was applied to Adam optimizer. The learning rate was monitored and reduced when the validation score plateaued.  $L_2$  regularization and early stopping methods were used to prevent over fitting. For evaluation of the segmentation work, mean dice coefficient (DSC) and mean intersection over union (IOU) were computed on the validation set.

## 4 RESULT

Test results of our model has been compared with the result of Pranet [2] below in table 1. The results taken from *Pranet* was also trained on *Kvasir* dataset. The metrics used were mean dice score, mean IOU and frame per second (FPS).

Model	Mean Dice	Mean IOU	FPS
Pranet	89.88	84.0	50
AttUNet(ours)	72.8	64.37	116.79

Table 1: Preliminary results based on validation and test set.

The *AttUNet* model produced a good validation result but did not generalized well as expected on unseen set of test data. However, the model was able to converge in 50 epochs. It also gave a very high FPS value in compare to *Pranet* which suggests that the adopted model can yield smooth real time results. There is room for improving the model's prediction by doing some hyperparameter search and adopting regularization techniques like weight averaging.

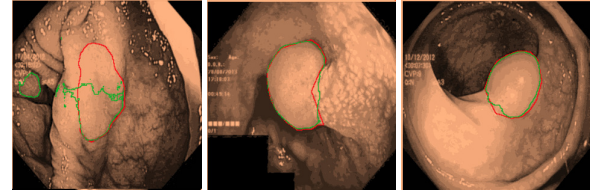


Figure 3: Segmentation result where green color indicates the ground truth outline and red color indicates the predicted outline.

## 5 CONCLUSION

In this work, the fine tuning of pretrained UNet model with attention mechanism has shown some promising results. This approach removes the requirement of an external object localization model and thus makes the model much simpler that yielded a high FPS value. The fine tuning of the pretrained model helped to converge faster without requirement of large number training examples. This indicates the importance and power of transferring prior knowledge from one domain to another. As reported in recent literature, attention mechanism have been a crucial factor in enhancing performance of various models. Lastly, we also note that this notion of knowledge transfer has been adopted successfully by meta-learning algorithms while learning from various tasks. This can serve well while solving problems in medical setting where there exist scarcity of labeled data and impact of different kinds of data shifts.

## 6 ACKNOWLEDGMENT

The research has benefited from the Experimental Infrastructure for Exploration of Exascale Computing (eX3), which is financially supported by the Research Council of Norway under contract 270053.

## REFERENCES

- [1] D. Bahdanau, K. Cho, Bengio, and Y. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
- [2] Fan Deng-Ping, Ji Ge-Peng, Zhou Tao, Chen Geng, Fu Huazhu, Shen Jianbing, and Shao Ling. 2020. PraNet: Parallel Reverse Attention Network for Polyp Segmentation. *IMICCAI* (2020).
- [3] Debesh Jha, Steven A. Hicks, Krister title=Medico Multimedia Task at MediaEval 2020:Automatic Polyp Segmentation Emanuelsen, Håvard D. Johansen, Dag Johansen, Thomas de Lange, Michael A. Riegler, and Pål Halvorsen. 2020. In *Proc. of MediaEval 2020 CEUR Workshop*.
- [4] D. Jha, P.H. Smedsrud, M.A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, and D. Johansen. 2019. Resunet++: An advanced architecture for medical image segmentation. *IEEE ISM* 9(2) (2019), 225–2255.
- [5] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. 2020. Kvasir-SEG: A Segmented Polyp Dataset. In *Proc. of International Conference on Multimedia Modeling (MMM)*. 451–462.
- [6] Mateusz buda. 2015. U-NET FOR BRAIN MRI. (2015). [https://pytorch.org/hub/mateusz buda\\_brain-segmentation-pytorch\\_unet/](https://pytorch.org/hub/mateusz buda_brain-segmentation-pytorch_unet/)
- [7] Luong Minh-Thang, Pham Hieu, and Manning Christopher D. 2015. Effective Approaches to Attention-based Neural Machine Translation. *arXiv:1508.04025* (2015). <https://arxiv.org/abs/1508.04025>
- [8] V. Mnih, N. Heess, A. Graves, and et al. 2014. Recurrent models of visual attention. *Advances in neural information processing systems* (2014).
- [9] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. 2018. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv:1804.03999* (2018).
- [10] Fischer P. Brox Ronneberger, O. 2015. U-net:Convolutional networks for biomedical image segmentation. *MICCAI* (2015).
- [11] J. Silva, A. Histace, O. Romain, X. Dray, and B. Granado. 2014. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery* 9(2) (2014), 283–293.
- [12] Gurudu S.R. Liang J Tajbakhsh, N. Automated polyp detection in colonoscopy. *IEEE* 35(2) (????), 630–644.
- [13] Marc Coram Varun Gulshan, Lily Peng. 2016. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* (2016). <https://jamanetwork.com/journals/jama/fullarticle/2588763>
- [14] Yan Kang Yingjie Zhao Qixun Qu Zhiqiong Wang, Ge Yu. 2012. Breast tumor detection in digital mammography based on extreme learning machine. (2012). <http://faculty.neu.edu.cn/bmie/wangzq/image/lunwen/17.pdf>