



# Transfer Learning in Polyp and Endoscopic Tool Segmentation from Colonoscopy Images

Nefeli Panagiota Tzavara<sup>1</sup> and Bjørn-Jostein Singstad<sup>2</sup>

1. *National Technical University of Athens, Athens, Greece*

2. *Oslo University Hospital, Oslo, Norway*

*E-mail any correspondence to: [b.j.singstad@fys.uio.no](mailto:b.j.singstad@fys.uio.no)*

## Abstract

Colorectal cancer is one of the deadliest and most widespread types of cancer in the world. Colonoscopy is the procedure used to detect and diagnose polyps from the colon, but today's detection rate shows a significant error rate that affects diagnosis and treatment. An automatic image segmentation algorithm may help doctors to improve the detection rate of pathological polyps in the colon. Furthermore, segmenting endoscopic tools in images taken during colonoscopy may contribute towards robotic assisted surgery. In this study, we used both pre-trained and not pre-trained segmentation models. We trained and validated both on two different data sets, containing images of polyps and endoscopic tools. Finally, we applied the models on two separate test sets. The best polyp model got a dice score = 0.857 and the test instrument model got a dice score = 0.948. Moreover, we found that pre-training of the models increased the performance when segmenting polyps and endoscopic tools.

**Keywords:** Polyp segmentation, Transfer learning, Hyperkvasir, Convolutional neural networks, MedAI challenge

## Introduction

Colorectal cancer (CRC) was the third most common and second most deadly cancer type worldwide in 2020 [1]. CRC is strongly associated with colorectal polyps, and colonoscopy is considered to be the best method for the detection of colorectal polyps [2, 3]. Studies have shown that between 6% and 27% of the colorectal polyps are missed by the clinicians during the colonoscopic examination [4]. On the other hand, artificial intelligence (AI) and image segmentation have shown to be useful in segmenting colorectal polyps [2, 3], and this may help the endoscopists to detect the polyps that otherwise are being overseen. Detection of colorectal polyps and endoscopic

tools may also play a role in the development of robotic-assisted surgical systems [5]. A recent study showed that pre-trained Convolutional Neural Networks (CNN) improved the performance in classifying colorectal polyps from colonoscopy images [6], but still it is not explored whether a pre-trained segmentation models will improve the performance of colorectal polyp segmentation. In this study, which is a part of a machine learning challenge [7], we aim to assess pre-trained and not pre-trained CNNs to detect polyps and endoscopic tools from colonoscopic images.

## Methods

Two models were developed as part of the challenge; one model to segment polyps in images and another to segment endoscopic tools in images. A CNN is a data-driven type of model and thus we had to train the model on some relevant data. The polyp model was trained on the Kvasir-SEG open data set consisting of 1000 images, containing one or more polyps [8], whereas the instrument model was trained on Kvasir-Instrument, which is another open data set consisting of 590 images, containing different endoscopic tools [5]. Both of the data sets also contained a corresponding annotated mask for each of the images, highlighting the polyps or endoscopic tools in the images.

**Data preprocessing:** The images and masks in the data sets vary in resolution, and thus, they had to be resized in order to be fed to the CNN models. We selected 256x256 pixels as the size of the input image and the predicted mask.

**Model architectures:** The model architectures were retrieved from a Python library; "Segmentation Models" [9], that contains different CNN architectures. This library provides models with both untrained and pre-

trained weights. Pre-trained weights are achieved by training on ImageNet [10]. To find the best fit for our data sets we tested the following architectures provided by the library: EfficientNet, MobileNet, SE-ResNet, Inception, ResNet and VGG. The results of these experiments are publicly available.<sup>1</sup>

**Augmentations:** Augmentations were applied on the training data in order to create a more versatile data set and achieve better generalization. We used nine different augmentation techniques: Random noise, gaussian blur, random rotation, image brightness, horizontal flip, vertical flip, random horizontal shift, random vertical shift and random zoom, for which an unique integer from 1 to 9 was assigned. For each epoch, the images and masks used to train the models, were given a random integer between zero and nine. The augmentation technique with the corresponding integer were used on the given image and mask. If the random integer was zero, no augmentation was applied.

**Model selection** 10-folded cross-validation on the development set were used to find the best model architecture and model parameters. The performance were measured using Dice similarity coefficient (DSC) and Intersection over Union (IoU) on the validation folds.

In the model selection phase, the learning rate was reduced during training, using a learning rate scheduler, which was set to lower the learning rate by a factor of ten when the IoU-score did not improve over three consecutive epochs.

#### *Clinical relevance and model transparency*

A polyp segmentation algorithm, like the one presented in this study, could probably be used as a decision tool for endoscopists. To make the segmentation tool more clinically relevant, and to streamline the work of the endoscopists, we developed a polyp counter algorithm. This algorithm detects the contours of the segmented polyps in the masks and counts objects. The purpose of this algorithm is to tell if or how many polyps there are in each image, so the doctors only need to look at the images with detected polyps and ignore the images without detected polyps. Moreover, the masks provided will highlight the polyps and improve the endoscopists focus on the abnormalities in the colonoscopy images. The polyp counter algorithm and the rest of the code developed in this project are publicly available on GitHub<sup>2</sup>.

## Results

**Model deployment:** From our experiments, we found that efficientnetb1 outperformed the other model architectures tested. Furthermore, we did experimental fine tuning of the hyperparameters and the settings which gave the highest mean DSC are shown in Table 1. Efficientnetb1

with the settings shown in Table 1 was finally used to train the models on the whole development set and applied on the test data which consisted of 300 images with endoscopic tools and 300 images with colorectal polyps. The predicted masks were used to participate in the MedAI challenge. In the final training procedure, the learning rate schedule was programmed to imitate the best learning rate schedule found during model selection.

Parameter	Polyp	Instrument
Model architecture	efficientnetb1	efficientnetb1
Pre-trained	Yes	Yes
Batch size	30	30
Epochs	20	35
Initial learning rate	0.001	0.001
Optimizer	Adam	Adam
Loss function	IoU	IoU

Table 1: Model architectures, parameters and hyperparameters used to train the final Polyp and Instrument model.

The segmentation performance of the best instrument and polyp models are summarized in Table 2.

Data set	Metric	Development set	Test set
Polyp	DSC	$0.874 \pm 0.011$	0.857
	IoU	$0.804 \pm 0.013$	0.800
Instrument	DSC	$0.937 \pm 0.015$	0.948
	IoU	$0.893 \pm 0.020$	0.911

Table 2: Dice similarity coefficient (DSC) and Intersection over Union (IoU) score achieved on both the polyp and instrument development sets and test sets. The scores on the development sets are achieved using 10-fold cross-validation.

The same models described in Table 1 and scored in Table 2, but without pre-training on ImageNet, were scored on the development sets using cross-validation. The model applied on the polyp data set achieved a DSC score of  $0.653 \pm 0.072$  and a IoU score of  $0.541 \pm 0.084$ . The model applied on the instrument data set achieved a DSC score of  $0.888 \pm 0.028$  and a IoU score of  $0.822 \pm 0.036$ .

## Conclusion

The results of this study show that the model which performed best on the development sets, according to our experiments, also generalized well to the MedAI test sets. Secondly, we found that pre-training the model on Imagenet significantly increased the performance on both the polyp and instrument development sets. These results may have implications for further work within the field of polyp segmentation, but also in other image segmentation tasks.

<sup>1</sup><https://app.neptune.ai/o/SSCP/org/HyperKvasir/experiments>

<sup>2</sup><https://github.com/ylefen/medai2021-polypixel>

## References

1. Xi Y and Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Translational Oncology* 2021; 14
2. Guo Y, Bernal J, and J. Matuszewski B. Polyp Segmentation with Fully Convolutional Deep Neural Networks—Extended Evaluation Study. *Journal of Imaging* 2020 Jul; 6:69
3. Jha D, Smedsrud PH, Johansen D, Lange T de, Johansen HD, Halvorsen P, and Riegler MA. A Comprehensive Study on Colorectal Polyp Segmentation with ResUNet++, Conditional Random Field and Test-Time Augmentation. 2020
4. Ahn SB, Han DS, Bae JH, Byun TJ, Kim JP, and Eun CS. The Miss Rate for Colorectal Adenoma Determined by Quality-Adjusted, Back-to-Back Colonoscopies. *Gut and liver* 2012 Jan; 6:64–70
5. Jha D, Ali S, Emanuelsen K, Hicks SA, Thambawita V, Garcia-Ceja E, Riegler MA, Lange T de, Schmidt PT, Johansen HD, Johansen D, and Halvorsen P. Kvasir-Instrument: Diagnostic and Therapeutic Tool Segmentation Dataset in Gastrointestinal Endoscopy. *MultiMedia Modeling*. 2021
6. Kim YJ, Bae JP, Chung JW, Park DK, Kim KG, and Kim YJ. New polyp image classification technique using transfer learning of network-in-network structure in endoscopic images. en. *Scientific Reports* 2021 Feb; 11. DOI: 10.1038/s41598-021-83199-9
7. Hicks S, Jha D, Thambawita V, Riegler M, Halvorsen P, Singstad B, Gaur S, Pettersen K, Goodwin M, Parasa S, and Lange T de. MedAI: Transparency in Medical Image Segmentation. *Nordic Machine Intelligence* 2021
8. Borgli H, Thambawita V, Smedsrud PH, Hicks S, Jha D, Eskeland SL, Randel KR, Pogorelov K, Lux M, Nguyen DTD, Johansen D, Griwodz C, Stensland HK, Garcia-Ceja E, Schmidt PT, Hammer HL, Riegler MA, Halvorsen P, and Lange T de. HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific Data* 2020 Aug; 7:283. DOI: 10.1038/s41597-020-00622-y
9. Yakubovskiy P. Segmentation Models. [https://github.com/qubvel/segmentation\\_models](https://github.com/qubvel/segmentation_models). 2019
10. Deng J, Dong W, Socher R, Li LJ, Li K, and Fei-Fei L. Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*. leee. 2009 :248–55