CS230

Semantic Segmentation of Colon Polyps in Colonscopy Images

Christian H. Nunez
Department of Physics
Stanford University
chnunez@stanford.edu

Abstract

It has been estimated that there is a 22% miss rate for colon polyps during routine colonoscopy examinations [1]. In this paper, we explore the task of semantic segmentation of colon polyps from colonoscopy video frames. Due to the small size of colonoscopy video datasets, data augmentation is applied synchronously to images and masks. These inputs are fed to one of two encoder-decoder networks: a U-Net or a SegNet. The tuned U-Net achieved a test dice coefficient of 0.48, which produces satisfactory segmentation masks for medium-sized, round polyps in the frame. Insights for improvement upon this model and with others are discussed.

1 Introduction

The American Cancer Society estimates that in 2019 alone, there have been 101,420 new cases of colon cancer and 44,180 new cases of rectal cancer in the United States [2]. Harvard Health reports that colon polyps can appear as “mushroom-shaped protrusions on the end of a stalk" or wide lumps on the colon wall. Colon polyps are predominantly benign; however, “adenomatous colon polyps" can cause mutations in the DNA of the colon lining and eventually lead to colon cancer [3]. In gastroenterology laboratories in medical centers, patients undergo a procedure known as a colonoscopy, where a tube equipped with a video camera (the colonoscope) is inserted into the rectum to view the colon wall. Very few implementations of automatic polyp detection exist today, and thus, this is an opportunity to learn from other biomedical segmentation projects to form a solution for this particular task.

This work is a binary semantic segmentation task, meaning that the goal is to correctly label each pixel of an image as "background" or "colon polyp." The input is a grayscale image taken from a colonoscopy video and the output is a binary mask prediction. This prediction is made through the use of one of two encoder-decoder fully convolutional networks: a U-Net [4] or a SegNet [5], which are both reliable segmentation architectures. The task of colon polyp segmentation is a quite difficult task, as Yao et. al. confirm, due to the irregularity of the size and the shape of colon polyps [6].

2 Dataset and Features

The CVC-ClinicDC dataset is a collection of 612 frames extracted from 29 different colonoscopy video sequences of colon polyps as viewed through a colonoscope [7]. As an important note, due to the fact that these 612 images are only collected from 29 sequences, adjacent images in sequences tend to look similar. Thus, it is crucial that the images in the train, validation, and test sets all come from different sequences. If a random shuffle were used instead, there would be a risk that images

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

in the training set may look nearly identical to some images in the test set. Thus, the test set would no longer composed of "unseen" data, compromising any conclusions that can be made about the model's ability to generalize. The train-validation-test split is the following: 435 images for training (sequences 9-29), 50 images for validation (sequences 1-2), 127 images for test (sequences 3-8). Xiao et. al. split the same dataset in roughly the same proportion for training vs. validation + test [8]. The resolution of the original images and the ground truth masks were RGB 384x288, but these were preprocessed to be 128x128 grayscale images. Even in the lower resolution grayscale images, the colon polyps were easily visible to the eye, thus, this transformation is appropriate. As expected, the colon polyps in different sequences come in a wide variety of shapes and sizes. In addition, the brightness of the video frames varies dramatically between sequences.

3 Related Works

3.1 End-to-End Deep Learning Approaches

Related work on the task of colon polyp segmentation falls under two categories: end-to-end deep learning approaches and computer-aided detection approaches. Implementations of colon polyp semantic segmentation via deep learning are quite sparse in the literature; here, I discuss two works that pull from the same dataset as this work and one that used a similar dataset.

Akbari et. al. attempted this same task of semantic segmentation of colon polyps using an FCN-8S, which is a fully convolutional network that "uses stages of convolution and pooling for creating dense feature maps for the input image" and thereby segments the candidate regions of the colon polyp [9]. Cleverly, this group used guided patch selection to enhance training and "Otsu thresholding" for post-processing. However, this work is quite concerning as only 300 images coming from 15 sequences were used for this network – 200 for training and 100 for test. Akbari et. al. cited that they chose the training and the test set images *randomly*, seemingly without regard for each image's sequence of origin, which, as discussed in the Dataset and Features section, is completely unadvisable for this dataset. Their cited 81% best dice score is questionable due to the likely contamination of the test set with images nearly identical to those in the training set. Xiao et. al. also attempted the task of semantic segmentation through the use of a novel architecture that combines Google's DeepLab_v3 network and Long Short-Term Memory (LSTM) networks [8]. This group used the full 612 images from the CVC-ClinicDB database for this task and achieved 93.21% mean intersection over union (mIoU). As a sequence model outfitted with LSTMs, combined with the fact that adjacent images in each sequence appear similar, a high mIoU is expected and was achieved. It appears that there was no sequence mixing between the training and the test sets. No data augmentation was mentioned in this implementation. Finally, Ribeiro et. al. experimented with a variety of CNNs (including the Fast-CNN, Alex Net CNN, and the GoogleLeNet CNN) for the task of automated classification of colon polyps from endoscopic images (therefore, the images were taken from within the patient's body) and demonstrated that CNNs are well-suited for the classification of colon polyps, especially when using in conjunction with transfer learning due to the small size of available colon polyp datasets [10].

3.2 Computer-Aided Detection (CAD) approaches

There are so few published deep learning approaches to this task that it is important to survey the current prevailing methods. These two following works attempt automatic colon polyp segmentation of colon polyps in computed tomography (CT) colonography images. CT scan images are produced by using X-rays to measure particular regions of the body and create a 2D "slice" image of the region. Thus, these images differ vastly in appearance to the CVC-ClinicDB optical colonoscopy images, but they are of colon polyps nonetheless. Yao et. al. created an automatic method of colon polyp segmentation in CT colonography images. The approach taken was a multi-step procedure in which CT scan images of the colon wall were fed into a surface-based filter and segmented into sub-images, and from there a series of automated (non-deep learning) techniques were used "polyp tissue" vs "non polyp tissue" (among other classes) and determine a polyp boundary. Nappi et. al. produced a similar CAD approach scheme for the detection of colon polyps in CT colonography images. We discuss this, because as Nappi et. al. note, CT colonography "could in time replace optical colonoscopy in the examination of the entire colon for large-scale screening applications" [11]. Future work could entail

semantic segmentation with deep learning of colon polyps in CT scan images instead of colonoscopy images.

4 Methods

4.1 Architectures

U-Net. This model is the primary architecture of this work, as it was used in the baseline and was later tuned extensively. The U-Net is a fully convolutional network used primarily for image segmentation tasks in cases where the training dataset is small [4]. Our architecture consists of four encoder steps and four decoder steps with "skip connections" to connect encoder levels with the equal resolution decoder levels to merge local and global information – a necessity for segmentation tasks.

SegNet. The SegNet, like the U-Net, is a convolutional encoder-decoder network. The main difference is non-linear upsampling is achieved by the decoder's use of the pooling indices of the corresponding encoder step (5 encoder/5 decoder steps in our model) [5]. The Seg-Net is more memory-efficient than FCNs, like the FCN-8S discused in Related Works.

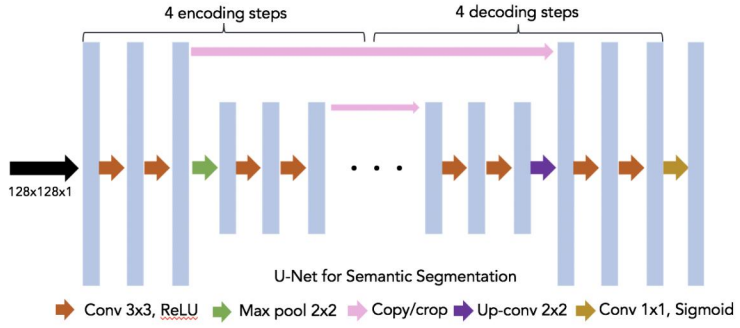

Figure 1: U-Net for Semantic Segmentation.

4.2 Baseline Review

In the U-Net baseline, the loss used for training was *dice coefficient loss*, which is the negative of the dice coefficient. The dice coefficient is effectively an intersection over union calculation. Thus, a dice coefficient close to 1 indicates perfect performance for general image segmentation networks. In particular, the dice coefficient D is the following:

$$D = \frac{2|A \cap B|}{|A| + |B|} \tag{1}$$

where A is the prediction mask and B is the ground truth mask. The training in the baseline was done without image augmentation of the 128x128 grayscale input images. Training resulted in 0.92 dice coefficient and 98.48% binary accuracy while the test set resulted in 0.285 dice coefficient and 80.40% binary accuracy. It is important to note that the polyps are very small in some images, so this model can achieve high binary accuracy with wildly incorrect segmentation predictions. Thus, the dice coefficient was chosen over binary accuracy as the optimizing metric (binary accuracy was included above to motivate this discussion). The disparity between the dice score of the train and test sets that there is a large variance problem – the solution: data augmentation. Nevertheless, it was notable that this architecture was complex enough to overfit the training set.

4.3 Learning Algorithms used in Experiments

The learning algorithm search space for the U-Net consists of Adam, RMSprop, and SGD with momentum.

1. SGD with momentum: Standard gradient descent with the alteration of using an exponentially weighted moving average (EWMA) of the gradients, and then using that average to update parameters.

2. RMSprop: Damps out oscillations of minibatch gradient descent by dividing the update to the parameters by the root of the EWMA of the squares of the gradients.
3. Adam: Combines the above two optimizers to adapt the parameter learning based on both the EWMA of the gradients and the EWMA of the square of the gradients.

5 Experiments, Results, and Discussion

We discuss several insights arising from hyperparameter turning. As previously discussed, the loss function is the *dice coefficient loss* and the primary metric used, in agreement with the literature in semantic segmentation, is the dice coefficient.

Learning rate. *Search space*: {0.0001, 0.001}. Training the U-Net from scratch on the non-augmented dataset with Adam and RMSprop, a batch size of 8, and at learning rate 0.001 resulted in dice coefficients of order 10^{-3} or less after 20 epochs. With Adam equipped with lower learning rate of 0.0001, in 20 epochs, the dice coefficient of the training set reached 0.6027. However, there is a variance problem: on the 127 test images, this model scored only 0.232 dice coefficient. Data augmentation is used to reduce the variance.

Data augmentation. Data augmentation was implemented with the Keras ImageDataGenerator synchronously to images and ground truth mask input pairs to combat the variance problem found in the baseline. Horizontal and vertical flips are justified due to the cylindrical geometry of the colon wall.

Implementation detail: ***pretraining***. With Adam, learning rate 0.0001 or 0.001, a model trained from scratch with augmented input would converge to low dice coefficients around order 10^{-3}. Thus, all future U-Net models were equipped with the pretrained weights from 20 epochs of training on the non-augmented training set optimized by Adam (learning rate = 0.0001).

Optimization algorithm. Adam, RMSprop, and SGD with momentum were tested on 60 epochs with data augmentation (lr = .0001 for Adam/RMSprop, momentum = 0.99 for SGD, otherwise keras.Optimizers standard values). All three algorithms displayed nearly identical training and validation dice coefficients at the end of the 60 epochs. Adam was chosen for further models.

Batch size. *Search space*: {1, 8, 16, 50}. Batch sizes 1, 8, 16, and 50 were tested, 8 was optimal on validation set.

The final U-Net hyperparameters chosen were: Adam optimization, .0001 learning rate, batchsize 8. Training was most sensitive to learning rate (too large of a learning rate caused poor dice scores) and least sensitive to optimization algorithm.

As the main architecture for this work was the U-Net, the time allotted for SegNet tuning was reduced, but there are several insights. Training with Adam(lr=0.001) with no data augmentation produced a variance problem. Data augmentation (of the same arguments as the U-Net) was applied to counter this, with pretrained weights from training without data augmentation.

5.1 Results

With the chosen hyperparameters, the tuned U-Net model was trained for 300 epochs and the SegNet for 102.

Summary of Results			
Model	Train dice_coef	Test dice_coef	Notes
Baseline	0.92	0.29	No data aug., Adam(lr=.001)
Tuned U-Net	0.56	**0.48**	Data aug., Adam(lr=.0001), batchsize=8
Tuned SegNet	0.33	0.23	Data aug., Adam(lr=.0001), batchsize=1

The best tuned U-Net decreased the variance problem dramatically and achieved a 0.48 dice coefficient on the test set. Below are three representative examples of prediction masks – I discuss the strengths and shortcomings of the tuned U-Net.

Upon analyzing the predictions for all 127 test set images, we draw several qualitative conclusions:

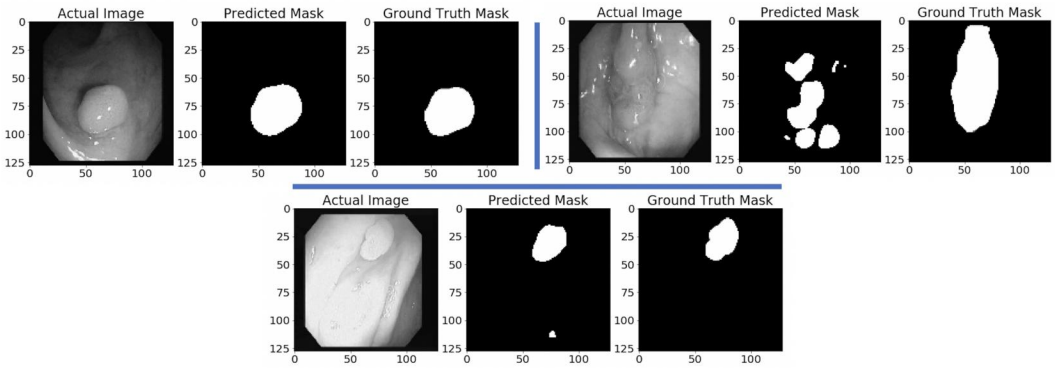

Figure 2: Prediction examples from the tuned U-Net.

1. Predictions are visually satisfactory for polyps that generate roughly circular masks, regardless of size.
2. Predictions are unsatisfactory for polyps that are long in shape and take up a large fraction of the image. For these types of polyps, the model often predicted masks made up of smaller, disconnected, roughly circular masks. In contrast, polyps that were excessively small in the frame were often missed.
3. Images with more brightness variation had more visually satisfactory predictions.

All 127 predictions can be found in the provided code repository. As compared to the U-Net, the tuned SegNet did not perform at the level of the U-Net, but likely more time for training would have allowed the SegNet to improve. We note that the SegNet seemed to localize the colon polyps well, but the prediction masks were much more granular than the U-Net's more solid, smooth masks.

6 Conclusion/Future Work

The U-Net model equipped with data augmentation, Adam optimization, learning rate 1e-3, and a batch size of 8 achieved the best performance on the test set (.48 dice score), which consisted of 127 images from 6 separate colonoscopy video sequences. All 127 image-mask-prediction triplets can be found in the provided GitHub repository (see under References).

6.1 Future work

1. Improve Current Models
   (a) U-Net: Use a larger encoder-decoder along with higher resolution, RGB images. Reasoning: Some polyps have a distinct color contrast compared to the colon wall. Also, attempt training with shears to help better generalize to non-round polyps. Finally, attempt guided patch selection, following Akbari et. al. [9].
   (b) SegNet: Due to time and resource constraints, the SegNet was not trained for an extensive period. More hyperparameter tuning and longer training will likely yield better results.
2. Explore New Models and Techniques
   (a) Implement transfer learning with FCNs trained on ImageNet, freezing early layers to retain the low level feature extractors.
   (b) Survey sequence models for real-time colon polyp segmentation.
3. Long-term
   (a) An exciting innovation with a sequence model that can process video in real time would be to develop a colonoscope that can automatically segment the video that the doctor would view.
   (b) Attempt to train/test on video sequences with polyps in only some of the frames in preparation for the above goal.

7 Contributions

Christian H. Nunez worked independently on this project, writing the code, performing the analysis, writing the final report, and constructing the poster. Special thanks to project teaching assistant Patrick Cho for helpful advice on the train-dev-test split, architecture choices, and data augmentation debugging.

References

[1] Jeroen C. van Rijn, Johannes B. Reitsma, Jaap Stoker, Patrick M. Bossuyt, Sander J van Deventer, and Evelien Dekker. Polyp miss rate determined by tandem colonoscopy: A systematic review. *American Journal of Gasotroentereology*, 101, 2006.

[2] Key statistics for colorectal cancer. 2019. American Cancer Society.

[3] Colon polyps. 2019. Harvard Health Publishing.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[5] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *Computer Vision and Pattern Recognition*, abs/1511.00561, 2015.

[6] Jianhua Yao, M. Miller, M. Franaszek, and R.M. Summers et. al. Colonic polyp segmentation in ct colonography-based on fuzzy clustering and deformable models. *IEEE Transactions on Medical Imaging*, 23, 2004.

[7] Bernal et. al. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics, 43, 99-111*, 2015.

[8] Wei-Ting Xiao, Li-Jen Chang, and Wei-Min Liu. Semantic segmentation of colorectal polyps with deeplab and lstm networks. *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, pages 1–2, 2018.

[9] Mojtaba Akbari, Majid Mohrekesh, Ebrahim Nasr-Esfahani, S.M. Reza Soroushmehr, Nader Karimi, Shadrokh Samavi, and Kayvan Najarian. Polyp segmentation in colonoscopy images using fully convolutional network. *CoRR*, 2018.

[10] Eduardo Ribeiro, Andreas Uhl, Georg Wimmer, and Michael Häfner. Exploring deep learning and transfer learning for colonic polyp classification. *Computational and Mathematical Methods in Medicine*, 2016, 2016.

[11] Janne Nappi, Abraham H. Dachman, Peter MacEneaney, and Hiroyuki Yoshida. Automated knowledge-guided segmentation of colonic walls for computeriaze detection of polyps in ct colonography. *Journal of Computer Assisted Tomography*, 26(4), 2002.

Code Repositories

1. U-Net: https://github.com/nikhilroxtomar/UNet-Segmentation-in-Keras-TensorFlow
2. SegNet: https://github.com/ykamikawa/tf-keras-SegNet

GitHub repository for this project: https://github.com/christianhnunez/ColonPolypSegmentation_CS230