Assignment 1 - Initiate

Ronald Fecso

October 20, 2022

The Kvasir SEG [JSR+20] dataset contains 1000 images of gastroinestinal polyps and their pixel-wise annotations, or segmentation masks. Polyps are a precursor to colon cancer which is one of the deadliest cancers in the world. Early detection of polyps is therefore crucial, but pixel-wise labels of polyps still remain few, due to the need for experts to hand-label these features. However, developing automatic methods of segmenting these images allows for better automatic detection of polyps in colonoscopy analyses. I will attempt to beat the results of two image segmentation models trained on this dataset.

This paper [ZSTL18] from Arizona State University proposes a U-Net with nested, dense skip pathways which try to reduce the loss of semantic information between the encoder and decoder network. It was trained on the Kvasir dataset and has state-of-the-art results on this task. Considering how many extra parameters there are in the network, I will try to test different architectures and hyperparameters to beat their segmentation results.

This paper [GCB21] from Bournemouth University proposes two encoder-decoder networks stacked back-to-back. The second network takes the original image and the prediction from the first network, which plays the role of an attention mechanism, as an input. This also resulted in state-of-the-art results on segmenting the Ksavir dataset. Considering this network was developed a year ago, I will consider changing the backbones of the encoder and decoder to something more recent and better performing.

Researchers from the University of Freiburg proposed a U-Net for segmenting medical images with a special focus on image augmentations [RFB15] . This is particularly helpful for the Kvasir dataset which is fairly small for a segmentation task. The methods proposed here may help in the pre-processing step by allowing the network to see more than 1000 images.

Below I describe the work-breakdown for each task.

1. Dataset collection: 5 hours. The database already exists. I simply need to upload it to an appropriate Python environment like Google Colab then figure out the best way to import and work with the images in Python.

2. Building networks: 10 hours. I need to import the networks from the researchers' Githubs into Python and learn how to work with them. Then, I need to decide what hyperparameters to test and how to tweak/add to the existing architectures.

3. Training and fine-tuning the networks: 15 hours. Training segmentation models is extremely time consuming so I will need to start training these models early and concurrently. They will probably not beat the state-of-the-art on the first try so I will need to re-train these models which will quickly increase the time needed for this step.

4. Building an application: 15 hours. I do not have a lot of experience developing web apps so this step might be more challenging. I need to allocate more time for this.

5. Writing the final report: 5 hours. Writing about something I will have been working for months should not be challenging.

6. Preparing the presentation: 5 hours. Organizing a presentation should also not be very difficult.

References

[GCB21] Adrian Galdran, Gustavo Carneiro, and Miguel A. Gonzá lez Ballester. Double encoder-decoder networks for gastrointestinal polyp segmentation. In *Pattern Recognition. ICPR International Workshops and Challenges*, pages 293–307. Springer International Publishing, 2021.

[JSR+20] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *International Conference on Multimedia Modeling*, pages 451–462. Springer, 2020.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[ZSTL18] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation, 2018.