Application of Deep Learning Models for Polyp Segmentation

Project Midterm Report

Abdurahman Ali Mohammed

Abstract

In the past few years, Deep learning has exhibited astonishing achievements in a way that it is competent with human beings. One of the areas that it gained a huge success is the field of computer vision. Tasks in computer vision range from simple classification task all the way to Object detection and semantic segmentation. Semantic segmentation is one of the hottest topics in computer vision as it can be used in various fields. This paper focuses on applying deep learning based semantic segmentation to detect polyps from colonoscopy images. Polyps are benign growths in the inner lining of the large intestine. Early discovery of these tissues is critical as it can helpful in preventing the progression of colon cancer. Three popular semantic segmentation models, namely: SegNet, FCN and UNet, are experimented with to identify the best performing model. As a measure of the performance of the models, Dice coefficient and Intersection over Union are used.

1 Introduction

A Polyp is an extra piece of tissue that grows inside the human body. It forms on the lining of the colon. Colons can develop into colon cancer which may be fatal when found in its later stages[1]. Hence, doctors need to take out any colon polyps found. These polyps can be found through different screening tests such as colonoscopy, colonography, stool tests and others. The early detection of polyps can be a huge factor in preventing colon cancer by enabling early measures to be taken[2].

In the world of computing, deep neural networks shown revolutionary performances in various artificial intelligence tasks such as Image classification[3], Image generation, and object detection[4]. These tasks intersect with various other fields to solve problems that would either take a lot of time to solve or are highly costly. One of these field is the medical field. With the intention to help medical professionals make accurate diagnosis, plan operations and evaluate prognopsis, automated systems can be highly beneficial[2].

One of the crucial tasks involved in the medical image processing task is Image segmentation. It can be applied in different computer aided diagnosis tasks such as brain-tumor detection, liver-tumor detection, lung segmentation and cardiac image segmentation[5].

Before deep learning came in to existence, early methods involved classifying hand engineered features independently[6]. Most deep-learning based computer vision tasks use Convolutional layers as their building blocks. Convolutional networks are commonly used for classification tasks where at the end of the network, a prediction for the given image is made[3]. Since using CNNs may help identifying that there is the object in the given image. However, it won't be much more precise than that.

This work focuses on experimenting with different deep learning-based image segmentation architectures and find out which architecture is best suited for Polyp segmentation task.

2 Related Work

Several attempts were made to perform image segmentation. The earliest attempts to detect polyps involved training of classifiers to remove non-polyp parts of a given image[7]. Other approaches used CRFs to infer both class segmentation and support relationships. Before the introduction of deep learning models, semantic segmentation mainly relied on conventional methods for image segmentation such as thresholding, clustering and region growing. It used hand-crafted low-level features to locate object boundaries in images. These earliest models tried to find inference by observing the dependencies between neighboring pixels[8].

After deep learning models start to get popularity, several attempts were made to create different architectures to perform pixel-wise classification. Of these attempts, the well known deep-learning architectures use the encoder decoder architecture[9]. In this architecture, the encoder part is used to extract features from given images and the decoder architecture is used to upsample and restore the extracted features back to the original image size.

3 Method

3.1 Dataset

For this work, the Kvasir-SEG[2] dataset was used. The dataset contains 1000 polyp images and their corresponding ground truth. Images in the dataset have size varying from 332x487 to 1920x1072 pixels. All 1000 images were manually annotated with the help of medical experts. The images in this dataset are organized in two directories: one for images and one for the masks.

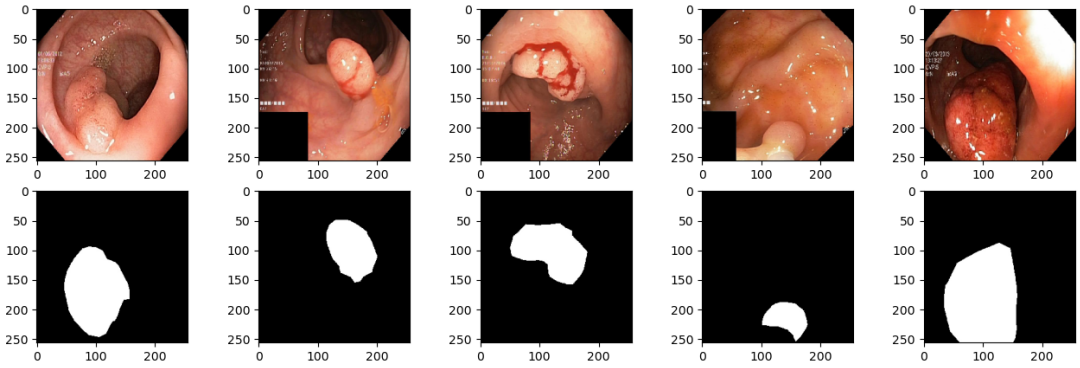

Figure 1: Sample images from the Kvasir-SEG dataset

3.2 Data preprocessing

The first thing to do in the image segmentation task was to pre-process the data. This stage started out by organizing the data into training, validation and testing datasets. The ratio that is used for train-valid-test split was 65-15-20 respectively. After preparing the datasets, the next task was to bring all the images into a consistent shape which was 256x256x3 which are *height* x *width* x *channels* respectively. The ground truth masks have shape 256x256x1.

After resizing the images, Normalization was applied to the pixel values so that each pixel has a value between 0 and 1. This was done due to the fact that neural networks process inputs using small weight values and inputs with large integer values can disrupt or slow down the learning process.

3.3 Model Architectures

Fully Convolutional Network (FCN)[10] was one of the first architectures that attempted to remove the fully connected layers. The idea behind it is to remove the dense layers of a classification network and add an upsampling mechanism to get pixel wise predictions. The upsampling technique used in FCN[10] is Transposed Convoltion which basically is regular convolutional layers with fractional strides. To improve the performance, the authors added skip connections from lower layers which helped recover some spatial information.

UNet[11] is another architecture that was introduced following FCN. What is different about it is that it uses multiple upsampling layers. It still uses skip connections and they are concatenated instead of being added. An input image is passed through the model and then it is followed by a couple of convolutional layers. The decoder part of the netwrok contains transposed convolution layers which is then concatenated with features from the encoder block. This helps keep features that sometimes might be lost due to the depth of the network. At the end of the network will be a sigmoid activation layer which will perform pixel-wise classification.

The other ground breaking semantic segmentation architecture is SegNet[9]. Instead of copying the encoder features as in FCN, indices from maxpooling are copied. This makes SegNet more memory efficient than FCN. The key component of the SegNet is the decoder network whcih consists of hierarchy of decoders one corresponding to each encoder. Then uses the max-pooling indices received from the corresponding encoder to perform non-linear upsampling of their input feature maps.

For the encoder part of the three architectures, any convolutional network can be used. To achieve better feature extraction with out slowing down the training process, a pre-trained VGG16[12] model was used as an encoder. To use VGG16 as an encoder for the segmentation task the fully connected layers were removed.

The models are implemented using Keras with Tensorflow as a backend. Techniques such as reducing the learning rate on plateu of validation loss with patience of 10 epochs was used. With early stopping applied (considering validation loss), the model was trained for 111 epochs with learning rate starting at $1e^{-3}$ and Adam optimizer. The batch size used for training was 32.

3.4 Evaluation metrics

There are two metrics recommended by [2] for the evaluation of models on the Kvasir-SEG dataset.

- **Dice coefficient** is a metric that compares the pixel-wise results between predicted segmentation and ground truth.

$$\text{Dice\_coef}(A, B) = \frac{2 * |A \cap B|}{|A| + |B|} \tag{1}$$

- **Intersection over union** is a metric that calculates the similarity between the predicted mask and the ground truth. It is also known as the Jaccard Index.

$$\text{IoU}(A, B) = \frac{A \cap B}{A \cup B} \tag{2}$$

4 Preliminary Results

Model	Dice_Coef	IoU
UNet	80.75 %	67.97 %
FCN		
SegNet		

Table 1: Performance of architectures using Dice and IoU metrics

Table 1 shows the performance of the 3 model implemented on a testing data from the Kvasir-SEG dataset. The UNet model implementation with VGG16 as an encoder demonstrated promisin results with a Dice score of 70.66% and IoU score of 55.10%. [1]

Figure 2 shows qualitative results of the models by comparing the predicted masks with the ground truth of the dataset used. The UNet model demonstrated a descent performance in predicting masks

[1] Results of FCN and SegNet will be included in the final report.

that are quiet close to the ground truth. However, looking at the second row, we can see that regions that are not labeled as polyp are also segmented.

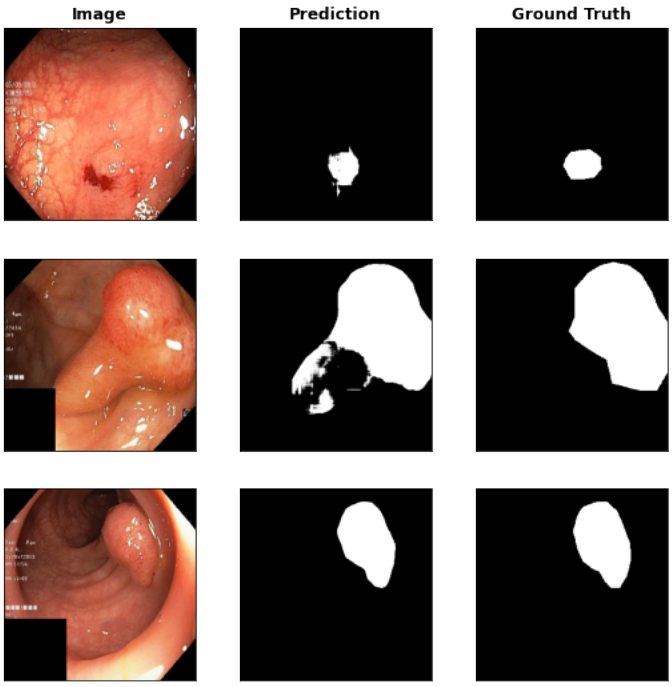

Figure 2: Predicted mask samples taken from UNet model and ground truth mask

5 Future plan

As part of future plans, Other architectures will be experimented along with different pre-trained networks. With enought experiments to fine tune the hyper parameters, this can help improve improve the performance of the models. In addition to that, ensemble of well-performing segmentation models can be used to get a better result in polyp segmentation. This will help leverage the strength of different upsampling techniques to build a stronger model.

6 References

[1] Michael B Huck and Jaime L Bohl. "Colonic Polyps: Diagnosis and Surveillance". In: (2016). DOI: 10.1055/s-0036-1584091. URL: http://dx.doi.org/.

[2] Debesh Jha et al. "Kvasir-SEG: A Segmented Polyp Dataset". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11962 LNCS (Nov. 2019), pp. 451–462. ISSN: 16113349. DOI: 10.48550/arxiv.1911.07069. URL: https://arxiv.org/abs/1911.07069v1.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F Pereira et al. Vol. 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[4] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2016-December (June 2015), pp. 779–788. ISSN: 10636919. DOI: 10.48550/arxiv.1506.02640. URL: https://arxiv.org/abs/1506.02640v5.

[5] Rongjian Zhao et al. "Rethinking dice loss for medical image segmentation". In: *Proceedings - IEEE International Conference on Data Mining, ICDM* 2020-November (Nov. 2020), pp. 851–860. ISSN: 15504786. DOI: 10.1109/ICDM50108.2020.00094.

[6] Risheng Wang et al. "Medical Image Segmentation Using Deep Learning: A Survey". In: *IET Image Processing* (Sept. 2020). DOI: 10.1049/ipr2.12419. URL: http://arxiv.org/abs/2009.13120%20http://dx.doi.org/10.1049/ipr2.12419.

[7] Nima Tajbakhsh, Suryakanth R. Gurudu, and Jianming Liang. "Automated polyp detection in colonoscopy videos using shape and context information". In: *IEEE Transactions on Medical Imaging* 35.2 (Feb. 2016), pp. 630–644. ISSN: 1558254X. DOI: 10.1109/TMI.2015.2487997.

[8] Irem Ulku and Erdem Akagunduz. "A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D images". In: *Applied Artificial Intelligence* (Dec. 2019), pp. 1–45. DOI: 10.1080/08839514.2022.2032924. URL: http://arxiv.org/abs/1912.10230%20http://dx.doi.org/10.1080/08839514.2022.2032924.

[9] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.12 (Nov. 2015), pp. 2481–2495. ISSN: 01628828. DOI: 10.48550/arxiv.1511.00561. URL: https://arxiv.org/abs/1511.00561v3.

[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully Convolutional Networks for Semantic Segmentation". In: (Nov. 2014). URL: http://arxiv.org/abs/1411.4038.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351 (May 2015), pp. 234–241. ISSN: 16113349. DOI: 10.48550/arxiv.1505.04597. URL: https://arxiv.org/abs/1505.04597v1.

[12] Karen Simonyan and Andrew Zisserman. *VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION*. Tech. rep. 2015. URL: http://www.robots.ox.ac.uk/.