Deep Learning Architectures for Medical Segmentation Tasks

Srinivas Natarajan (1224944409), Sai Pranav Tavva (1225344341),
Tanushi Ahuja (1225475680), Pranavi Addagatla (1225696667),
Shivani Yerram (1225766373)

December 8, 2022

Abstract

The importance of early detection localization of polyps (precursors to colon cancer) cannot be understated. The error-prone nature of the manual screening process for such abnormalities necessitates an automated system that can help pinpoint the location of anomalies with ease. In this project, we aim to exploit deep learning models that specifically employ Convolutional Neural Networks (CNNs) and their variants for their state-of-the-art performance in image identification tasks. The objective is to build a model that can be generalized to detect similar anomalies in such medical procedures and tag them reliably.

1 Introduction

In recent years, medical segmentation has played a vital role in analyzing diseases, giving clinical diagnoses, and improving treatment. It is used for medical diagnoses of breast cancer, tumour detection, pneumothorax detection from X-rays etc. Given that Colorectal cancer is one of the fastest growing types of cancer and its primary method of detection is by collecting polyp samples through colonoscopies. This method can be a little hit or miss due to the difficulty in detecting polyps in the body. It would be effective to have a computer-aided system to identify polyps and lesions during a medical procedure, especially when used in conjunction with a medical expert. The most commonly used methodology in medical segmentation problems involves Convolutional neural networks and their variations. This is due to their efficiency in processing larger images by extracting smaller features from them. Studies have found architectures such as UNet and ResUNet have shown great promise in tackling this problem but still require some tuning due to the contextual nature of the problem. However, CNN-based models in general show limitations for explicit long-range relations and they might exhibit unstable performance, unlike transformers. This is why more recent advances in the field have been implementations of a fusion between transformer architectures, taking their encoderdecoder framework and implementing them with CNNs. This project is based on the pattern recognition and the neural networks segment of our syllabus and will implement the concepts of feature selection, clustering and dimensionality reduction as covered in class.

2 Motivation

Colorectal Cancer (CRC) has the third highest mortality rate among cancers with a five year survival rate of 68%. Given the severity of this disease, it is important to identify precursors like polyps as early as possible to increase the survival rate of patients. The main way to identify these precursors is through endoscopic procedures but they are often hard to spot even to experienced doctors. This is why we felt it was crucial to develop systems that can aid doctors in recognizing these obscure growths and mark

them out for further examination.

3 Problem Description

Polyps are abnormal tissue growths from the mucous membrane found in the gastrointestinal tract. They are a precursor to more serious colon diseases and is an important sign that is checked through endoscopy procedures. They can be hard to detect as they are usually small in size and only have very minor pigmentation differences from the rest of the tract. So our goal is, given a set of images of the gastrointestinal tract (GI), we must accurately isolate these growths and mark their regions to aid doctors in the identification process.

4 Methodology

Datasets: To train the classification models, we must first find appropriate images of colonoscopy procedures for the detection of polyps. We selected the Clinic CVC which has 612 still images from 29 different sequences and the Kvasir-SEG data sets consisting of 1000 images of GI endoscopies. Each of these data sets contain the main image as well as ground truth masks for a supervised approach.

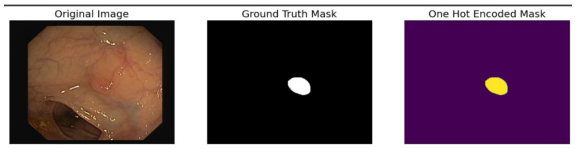

Figure 4.1: Polyp Image and ground truth

Preprocessing: The images in the dataset come in varying sizes and thus need to be standardized. We chose a resolution of 288x384 to enable an even padding of size 32. We keep them in the RGB format as this information contributed to better results. But we standardize the 0-255 values of the pixels into a 0-1 range. As the number of examples we have are too little to achieve good results, we augment our training data. We do this by implementing a data augmentation function that flips the images both horizontally and vertically, randomly rotate and crop the image and vary the brightness.

We chose five models for comparison, all of which are based on a auto encoder architecture. We use a variety of feature extraction networks based on the image net architecture and weights. The base architecture we use is the U-net model, a base standard in the field of medical segmentation. It gets its name from the U shaped architecture consisting of a Contraction and an Expansion phase. The images are first down-sized using convolution layers, retaining important information which is later fed into the expansion layer to reconstruct an image with just the essential information.

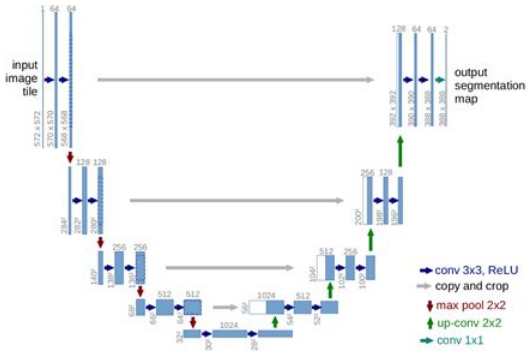

Figure 4.2: UNet Architecture

This symmetric architecture consists of two main components- contraction and expansion. The contraction section is a classic CNN architecture consisting of two 3x3 convolution layers, a ReLU and 2x2 max pooling unit. At each stride of the max pooling operation, we double the number of feature channels so in the expansion section, we have 2x2 up-convolution layer that halves the number of feature channels. Then, we have concatenation, two 3x3 convolution layers. Each of these layers are followed by a ReLU. Finally, we have 1x1 convolution layer to map each feature vector to a suitable number of classes.

UNet++ is a convolutional neural network (CNN) architecture for image segmentation. It is an improved version of the original U-Net

architecture, which was designed for biomedical image segmentation. UNet++ is designed to better capture fine details and improve the performance of the model on a variety of image segmentation tasks. It achieves this by using a hierarchical, nested U-shaped architecture, where the skip connections between different levels of the network are carefully designed to allow for information flow between different scales of the input image. This allows the network to make more accurate predictions at the pixel level, leading to improved performance on a wide range of image segmentation tasks.

We also have another feature extractor known as a Feature Pyramid Network, or FPN, produces proportionally scaled feature maps at several levels in a completely convolutional manner from a single-scale image of any size. FPN consists of a bottom-up and a top-down pathway. Bottom-up Pathway: The feedforward computation of the backbone ConvNet is the bottom-up pathway. The final layer of each stage's output will serve as the reference set of feature maps for lateral connection-enhanced top-down enrichment. By upsampling geographically coarser but semantically stronger feature mappings from higher pyramid levels, the top-down approach creates the illusion of greater resolution features; the spatial resolution is upsampled by a factor of 2. Each lateral connection combines feature maps from the top-down and bottom-up pathways that are the same spatial size.

The fourth model is Pyramid Attention Networks which introduce two important modules, the Global Attention Upsample (GAU) and Feature Pyramid Attention (FPA). The Feature Pyramid Attention (FPA) uses a U shaped pyramid architecture to extract information using 3x3, 5x5 and 7x7 filters. The pyramid structure incorporates information extracted step by step without the additional burden of using larger filters as the resolution of the feature map is already small. The original features are multiplied by the pyramid features after passing it through a 1x1 convolution. Global Attention Upsample (GAU) performs global average pooling to provide global context. These features are passed through a 1x1 convolution along with batch normalization and the ReLU activation for non-linearity. These weighted low level features are added along with the high level features to provide more information.

The final model is DeepLab created by Google. It solves the problem of information loss as we downsample images through convolutions. It does this through Atrous Convolution layers (Dilation CNN) combined with Spatial Pyramid Pooling modules. It maintains features by using a technique called PointRend enhancement that employs subdivision algorithms to upscale the low resolution image. This increase the efficiency of the upscaling pyramid architecture.

5 Results

We compiled our results into a tabular format to compare the performances of the models. We analyse their IoU scores and their Dice Scores as seen below.

Table I
Performance Comparisons of Models

Model	IoU Score	Dice Loss
UNet	0.9684	0.0475
UNet++	0.9756	0.1250
FPN	0.8920	0.0574
PAN	0.9526	0.0326
DeepLab v3	0.9716	0.0500

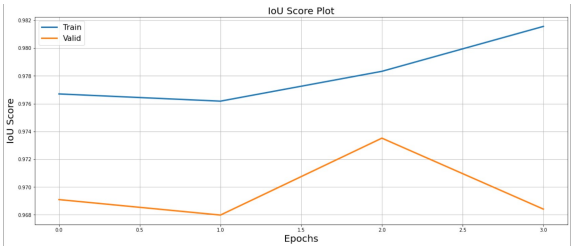

Figure 5.1: UNet IoU plot

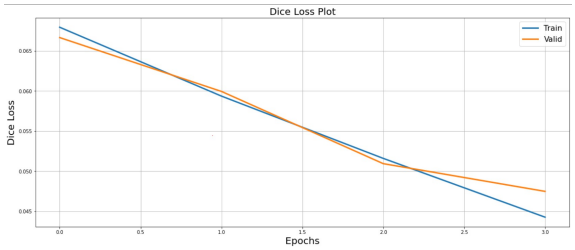

Figure 5.2: UNet Dice plot

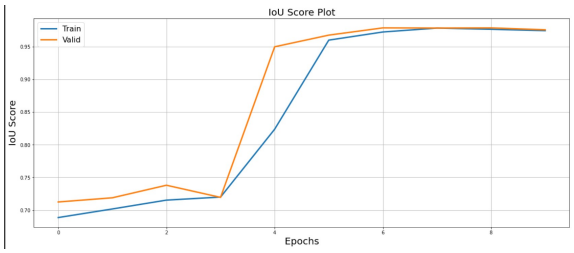

Figure 5.3: UNet++ IoU plot

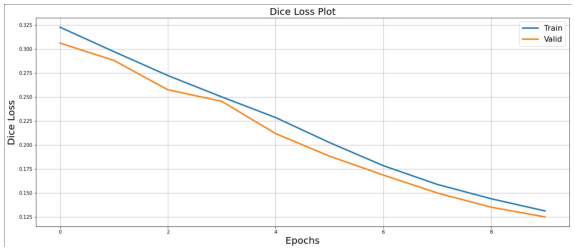

Figure 5.4: UNet++ Dice plot

From the IoU and Dice plots, we can observe that UNet++ performs better than UNet due to the reconfiguration and additional design of skip pathways. Deep Supervision also helps the model to converge faster by providing companion objective functions that act as a performance boost for the convergence.

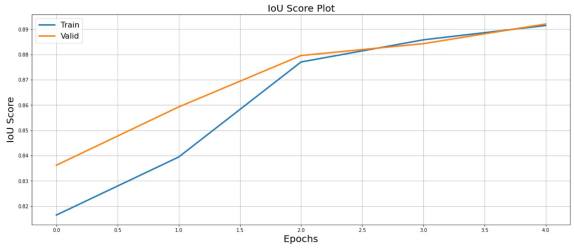

Figure 5.5: FPN IoU plot

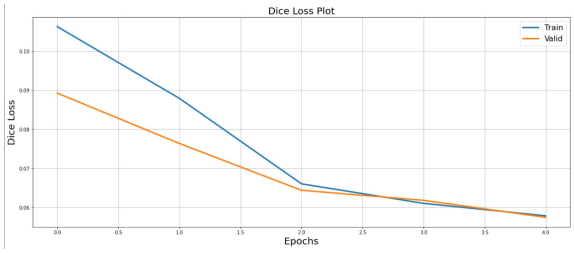

Figure 5.6: FPN Dice plot

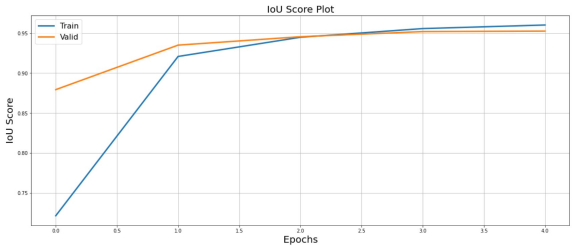

Figure 5.7: PAN IoU plot

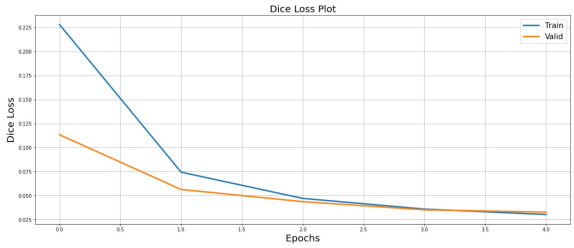

Figure 5.8: PAN Dice plot

Both Feature Pyramid Networks and Pyramid Attention Networks are feature extraction models known for capturing long range contextual information from features that exist across multiple, contrasting levels. The plots show similarity when it comes to the time and complexity they consume for training the models. The upsampling provides more spatial information in both models.

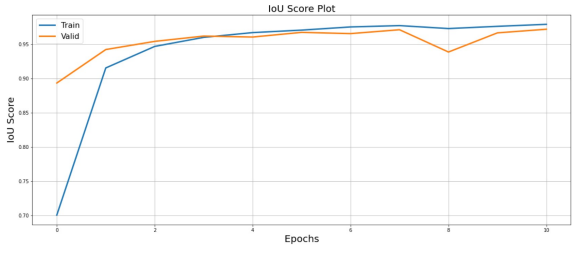

Figure 5.9: DeepLab IoU plot

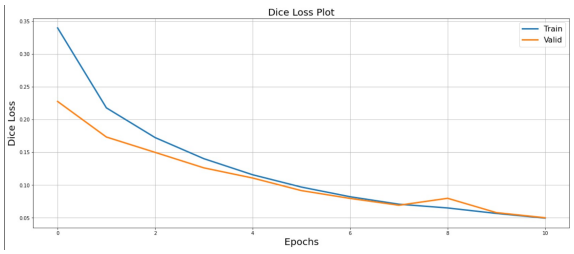

Figure 5.10: DeepLab Dice plot

With the help of atrous convolutional layers and a more precise upscaling method based on the Point Rend technique developed by Google, we see that the DeepLabv3 architecture performs among the best in terms on IoU score. It is comparable with the UNet++ model but has a superior dice loss metric. This is why we chose this as our final model for polyp segmentation.

6 Related Work

Over the past two decades, research on automated polyp detection has been vigorous, and a lot of work has been done to develop effective techniques and algorithms. Earlier studies employed handmade descriptors-based feature learning and paid special attention to polyp color and texture. Methods for computer-aided segmentation, localization, and detection can potentially make colonoscopy procedures better. Even though numerous techniques have been developed to deal with the automatic detection and segmentation of polyps, the challenge of bench marking cutting-edge techniques is still unresolved[1]. In [1] we find that the authors benchmark several state-of-the-art methods using Kvasir-SEG an open-access dataset of colonoscopy images for polyp detection, localisation, and segmentation evaluating both method accuracy and speed. Sizes of gastric polyps vary. Small polyps are challenging to distinguish from the background, which makes polyp detection challenging.To solve this problem authors of paper [2] feature extraction and fusion module. This approach outperforms competing approaches in the detection of small polyps because it can combine semantic data from high-level feature maps with low-level feature maps. Early detection of CRC by screening procedures like colonoscopy and sigmoidoscopy is essential for boosting the survival rate.Addressing this issue the authors of [3] present a systematic study to facilitate the development of deep learning models for video polyp segmentation (VPS). In [4], the authors use convolutional neural networks (CNN) and a succinct model called the Gastric Precancerous Disease Network to classify three types of gastric precancerous disease (GPD), namely polyp, erosion, and ulcer (GPDNet).

7 Conclusion

While a precise segmentation mask may not be critical in natural images, even marginal segmentation errors in medical images can lead to poor user experience in clinical settings. Endoscopic assessment of severity and sub-classification of different findings may also vary from one doctor to another. Accurate grading of diseases are important since it may influence decision-making on treatment and follow-up. Our metrics thus emphasize a greater penalty for misclassifications of polyp boundaries. While all our models performed well, we can see the improvements in more recent developments through the introduction of better upscaling algorithms and computations. They are also relatively faster due to the used of atrous convoluional layers which can mimic the wider perspective of larger kernels while avoiding the calculation penalty and the processing requirements that comes with them.

8 References

[1] Real-Time Polyp Detection, Localization and Segmentation in Colonoscopy Using Deep Learning," in IEEE Access, vol. 9, pp. 40496-40510, 2021

[2] Cao C, Wang R, Yu Y, zhang H, Yu Y, Sun C (2021) Gastric polyp detection in gastroscopic images using deep neural network. PLoS ONE 16(4): e0250632.

[3] M. Akbari et al., "Polyp Segmentation in

Colonoscopy Images Using Fully Convolutional Network," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 69-72
[4] Zhang X, Hu W, Chen F, Liu J, Yang Y, Wang L, et al. Gastric precancerous diseases classification using CNN with a concise model. PloS One. 2017; 12(9):e0185508. https://doi.org/10.1371/journal.
pone.0185508 PMID: 28950010