Building a Polyps Segmentation model using customized U-Net with ResNet-152 Backbone

Nguyen Cao Ky

HUST, July 21st, 2023

Abstract

The Sun* Internship Program 2023 commenced on June 15th, offering aspiring interns the opportunity to tackle the challenge of Polyp Segmentation and subsequently deploy the model through Streamlit. The dataset for this project is readily available for download on GitHub. Due to some reasons, the writer chose to adopt a simplified U-Net architecture, utilizing ResNet-152 as the backbone. This report focuses on the project organization, research methodology, technical reasons behind choosing loss function, evaluation metrics and insights drawn from results.

I Introduction

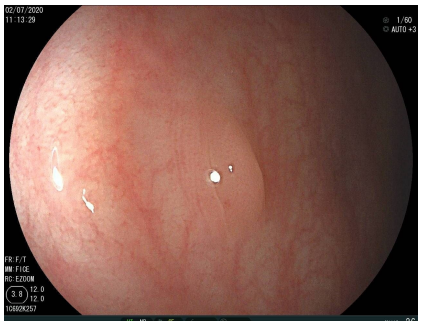

Figure 1: Illustration of a scarcely discernible polyp

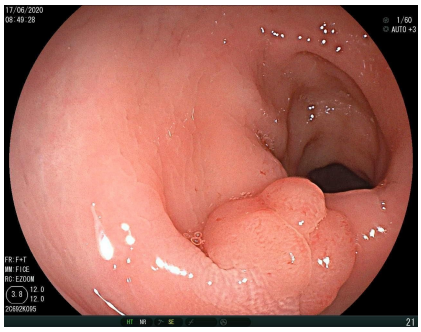

Figure 2: An Incipient Polyp in Medical Imaging

Recent advancements in Computer Vision have significantly transformed the field of medicine, particularly in Medical Imaging. Modern Image Segmentation models have played a crucial role in relieving the burden on medical systems, notably in tasks like polyp diagnosis and tumor detection. Contests on Medical Imaging are held worldwide annually, giving compelling evidence for the necessity of Computer Vision in the field.

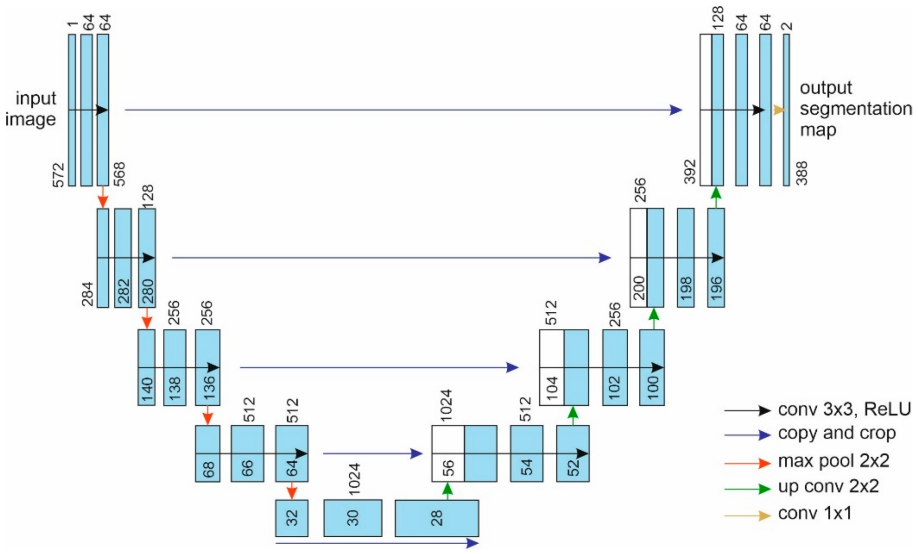

Figure 3: U-Net Architecture

While exploring the dataset, we encountered some significant challenges that need to be addressed. First, polyps tend to blend with the background, making it difficult for the model to identify their edges accurately. Second, distinguishing polyps from surrounding tissue poses a considerable challenge due to their similarity to human tissue.

II Methodology

II.1 U-Net architecure

The U-Net architecture consists of an Encoder (contracting path) that reduces image dimensions and augments feature maps, a neck block (bottleneck) serving as a feature fusion module, and a Decoder (expansive path) for upsampling and generating pixel-wise segmentation predictions. The Encoder captures abstract features, the neck block combines semantic and spatial information, and the Decoder restores original spatial dimensions. U-Net excels in image segmentation, especially in biomedical applications.

In each stage of upsampling, the resulting feature map is concatenated with the corresponding down-sampling block in the Encoder, creating what is commonly referred to as a "skip connection." The implementation of skip connection is to reintroduce the fine-grained spatial information that would have otherwise been lost due to down-sampling. The skip connections foster multi-scale representations, enabling the U-Net to effectively combine both high-level semantic knowledge and fine-grained spatial details. This dynamic interplay enhances

the model's ability to precisely delineate object boundaries and produce accurate segmentation masks, especially in scenarios with complex and intricate structures.

The final stage of the U-Net architecture involves passing the last feature map through a 1×1 Convolution Layer, serving multiple critical purposes:

1. **Dimensionality Reduction:** The conv1x1 layer reduces the number of channels or feature maps, thereby enhancing computational efficiency and conserving memory. By employing a 1×1 kernel, the layer enables controlled adjustment of the output channels.

2. **Feature Combination:** The conv1x1 layer assumes a pivotal role in feature fusion within the U-Net's bottleneck. It facilitates the seamless amalgamation of high-level semantic information from the contracting path with the spatial details from the expansive path.

3. **Non-linearity:** Although individual convolutions with a 1×1 kernel are inherently linear, the incorporation of non-linear activation functions (e.g., ReLU) introduces non-linearity to the network. This fosters the learning of intricate relationships and representations within the data.

4. **Adaptive Weighting:** Leveraging the conv1x1 layer, the network can assign learnable weights to distinct channels, empowering it to assign varying degrees of importance to different features during the information fusion process.

In summary, the 1×1 Convolution Layer at the end of the U-Net plays a crucial role in optimizing performance by reducing dimensionality, merging features, introducing non-linearity, and enabling adaptive weighting for enhanced feature representation and accurate segmentation output.

II.2 Metrics

Metrics are used to monitor and measure model performance. In the training process, the metrics used are the Dice score and the IoU score. These metrics are particularly valuable in tasks like image segmentation and medical imaging.

1. Dice Score (also known as F1-Score):

   The Dice score is a common evaluation metric used in binary classification tasks, particularly in image segmentation and medical imaging applications. It assesses the similarity between the predicted and ground truth segmentation masks.

$$\text{Dice} = \frac{2 \cdot |A \cap B|}{|A| + |B|} \tag{1}$$

2. IoU Score:

   The IoU score is calculated as the ratio of the intersection of the predicted and ground truth masks to the union of these masks.

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

   Both scores range from 0 to 1, where a value of 1 indicates a perfect match between the predicted and ground truth masks, and 0 represents no overlap between the two masks.

III Results

Dataset	**IoU**	**Dice**
Train	0.7177979924462058	0.8938686468384482
Test	0.7094497203826904	0.897873063882192

The model exhibited consistent performance on both the train and test datasets, with IoU scores of approximately 0.71 and Dice scores around 0.89. This strong resemblance in evaluation metrics indicates effective generalization, as the model accurately segmented objects in previously unseen data. As a result, we can confidently conclude that the model is well-suited for the image segmentation task and demonstrates robust performance.

Given the model's successful generalization, it holds promise for real-world applications. However, to ensure broader reliability, further validation on diverse datasets is recommended. This additional testing will provide valuable insights into the model's performance in varying scenarios and enhance its suitability for practical use across a range of image segmentation applications.

IV References

1. U-Net architecture [Image]. Retrieved from Uni Freiburg. (n.d). https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/u-net-architecture.png