Generating data for polyp semantic segmentation

Mentors: Tan Cong Nguyen & Thao Dao

Members: Ly Phan & Khue Luu

DESCRIPTION

Medical data is crucial for developing machine learning models to improve diagnosis and treatment, but its use is hindered by challenges such as consent and security, time-consuming labeling, and limited data volume. This project aims to overcome these obstacles by generating new data from real medical images using SinGAN-Seg, which will significantly expand the dataset. By doing so, the project seeks to enhance the performance of machine learning models. Specifically, a U-Net model will be trained on three datasets: the original set of real polyp images, the newly generated images, and a combined set of both. The goal is to achieve more than a 2% improvement in the Dice score with the combined dataset compared to the original. This improvement will enhance the accuracy and generalizability of polyp detection models, thereby contributing to better diagnostic tools in the medical field. In our experiment, the model trained on the combined dataset attained the highest Dice score of 0.89, outperforming all other models.

LINKS

- Presentation Slides
- GitHub Repository