TransResNet: Integrating the Strengths of ViTs and CNNs for High-Resolution Medical Image Segmentation via Feature Grafting

Muhammad Hamza Sharif, Dmitry Demidov, Asif Hanif, Mohammad Yaqub, Min Xu

Introduction

Background

- High-resolution images have rich semantic information that can improve the diagnostic capability of the underlying method.

Challenges

- Existing DL methods for medical image segmentation (IS) are designed for input images having small spatial dimensions and perform poorly on high-resolution images due to discrepancies between sampling depth and receptive field size.

- CNN-based (IS) methods cannot capture global context details due to narrow and fixed receptive fields, while transformers-based (IS) methods are computationally prohibitive and often downsample the input image before processing.

Solution

- We need a combined approach that captures rich local and global information without increasing the computational complexity associated with high-resolution images and gives accurate segmentation results.

Method

- We introduce TransResNet using two encoder modules: One is CNN based for extracting local feature details, other is transformer based for extracting the underlying features.

- We introduce Cross Grafting Module (CGM) to combine the feature maps from both encoder branches. CGM generates grafted features enriched in both local and global semantic cues.

- We design our decoder in a staggered manner, which first receives the flow of features from the transformer branch, followed by a cross grafting module, and finally CNN branch.

Experimental Results

We evaluate our method on ten datasets for three segmentation tasks:

- Skin Lesion Segmentation (2 datasets)
- Retinal Vessel Segmentation (3 datasets)
- Polyp Segmentation (5 datasets)

Analysis

This research work is supported by the Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI). For more details, please contact: muhammad.sharif@mbzuai.ac.ae or friend me via LinkedIn.