Master Thesis

Multi-scale Transformer Model for Semantic Segmentation

Background

Despite the great success of CNN networks on semantic segmentation tasks, recent studies reveal that CNN networks usually perform the recognition task using the local texture information rather than a shape or geometrical information. The locality nature of the convolution operation (even with atrous technique) limits the strength of these networks in capturing the long-range contextual dependency. To tackle this problem, motivated by the significant success of transformers in NLP era, a line of research is perusing the applications of transformers in computer vision. Transformers, unlike regular CNNs, are not only capable of building global contextual representations, but also leverage to the versatile local information. They proved to have the power to overcome the entailment of modeling long-range iterations and spatial dependencies, which are arguably crucial for accurate performance in many challenging contexts such as semantic segmentation. However, the pure transformer models (e.g. vision transformer [1]) usually suffer from a poor representation of local information and are less precise in connecting the global and local information. The articles [2-4] are among the pioneering works proposed to jointly model local and global representation in an end-to-end manner. Although these methods perform well on the challenging semantic segmentation task, they usually suffer from weak local representation and consequently lack to capture multi-scale representation. In this research work, we aim to model the multi-scale representation within the Transformer model. Our objective is to model the interaction among the pyramid representation and incorporate an attention mechanism to adaptively recalibrate the feature representation. We will validate our findings on medical image datasets.

Tasks

- Literature review
- Implement and evaluate the baseline methods
- Investigate the solution for the literature performance
- Evaluate the performance in different settings
Your profile

- Student of RWTH Aachen with Faculty 6
- Strong programming skills (Python)
- Knowledge in computer vision and deep learning
- Experience in deep learning framework (Pytorch)
- Strong writing skill

Our Offer

Our institute features an ultra-modern computer infrastructure, including a remotely accessible cluster for training the deep learning networks with dozens of GPUs. Throughout the thesis period, you will be supervised with regular meetings and guidance.

References


If you are interested, please send a short email to azad@lfb.rwth-aachen.de with your resume and current transcript.