Lehrstuhl für Bildverarbeitung

RWTH AACHEN UNIVERSITY

Master Thesis

Multi-scale Transformer Model for Semantic Segmentation

Univ.-Prof. Dr.-Ing. Dorit Merhof
Lehrstuhlinhaberin

Reza Azad, M.Sc.
Wissenschaftlicher Mitarbeiter

Raum 104
Kopernikusstr. 16
52074 Aachen
GERMANY

Telefon: +49 241 80 22906
Fax: +49 241 80-22200

azad@lfb.rwth-aachen.de

03.11.2022

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 peruseing 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 iteractions 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 muli-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 litreature liperformance
- Evaluate the performane in different settings

- Write a paper (we will completely guide)

Your profile

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

Our Offer

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

References

1. Dosovitskiy, Alexey and et. al. "An image is worth 16x16 words: Transformers for image recognition at scale", arXiv preprint arXiv:2010.11929, 2020.
2. Chen, Jieneng, et al. "Transunet: Transformers make strong encoders for medical image segmentation." arXiv preprint arXiv:2102.04306 (2021).
3. Hatamizadeh, Ali, et al. "Unetr: Transformers for 3d medical image segmentation." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.
4. Zhang, Yundong, Huiye Liu, and Qiang Hu. "Transfuse: Fusing transformers and cnns for medical image segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2021

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