Lehrstuhl für Bildverarbeitung

RWTH AACHEN UNIVERSITY

Master Thesis

3D 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

22.02.2022

Background

The success of Transformers in Natural Language Processing (NLP), a machine translation task, recently made its steps to the computer vision field, namely Vision Transformers (ViT) [1]. Unlike regular CNNs, transformers can build global contextual representations and leverage versatile local information. During the last two years, tremendous cooperation with ViTs done in diverse vision tasks to replace the widely used convolution operations with the newborn Transformer pipeline due to its ability to learn the long-range contextual dependencies. In addition recent studies such as Swin Transformer [2] and Efficient Transformer [3] utilized to address the lack of hierarchical representation (due to the fixed-size patch embedding process) and the high computational burden (due to the quadratic complexity of Transformers). In medical image analysis, utilizing the Transformers in dense prediction tasks, e.g., segmentation, synthesizing, etc., successfully consolidates its footprints. MISSFormer [4], Swin-UNet [5], and TransUNet [6] are the signatures of the effectiveness of Transformers in the 2D medical image segmentation domain. Medical imaging plays a pivotal role in modern healthcare by utilizing in vivo examination of human body pathology. Nonetheless, due to the nature of medical imaging data and the representative power of 3D data, still not enough effort was made into this concept to model the inter-slice affinities. UNETR [7] encodes the 3D medical data to a 1D sequence for feeding a simple ViT pipeline, D-Former [8] proposes a dilated Transformer module to increase the scope of information interaction among 3D patches. In this research work, we aim to model the 3D relevancy of medical data across multiple slices in a hierarchical concept with a pure transformer model. We will validate our findings on medical image datasets.

Tasks

- Literature review
- Implement and evaluate the baseline methods
- Investigate the solution for the litreature limitations
- 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. Liu, Ze, et al. "Swin transformer: Hierarchical vision transformer using shifted windows." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
3. Wang, Zhendong, et al. "Uformer: A general u-shaped transformer for image restoration." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
4. Huang, Xiaohong, et al. "Missformer: An effective medical image segmentation transformer." arXiv preprint arXiv:2109.07162 (2021).
5. Cao, Hu, et al. "Swin-unet: Unet-like pure transformer for medical image segmentation." arXiv preprint arXiv:2105.05537 (2021).
6. Chen, Jieneng, et al. "Transunet: Transformers make strong encoders for medical image segmentation." arXiv preprint arXiv:2102.04306 (2021).
7. Hatamizadeh, Ali, et al. "Unetr: Transformers for 3d medical image segmentation." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.
8. Wu, Yixuan, et al. "D-former: A u-shaped dilated transformer for 3d medical image segmentation." Neural Computing and Applications (2022): 1-14.

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