Can a transformer architecture match convolutional neural networks for segmentation of anatomic structures in 3D computed tomography?

Gonçalo Almeida[a], João Manuel R.S. Tavares[b]

1. Introduction

Image segmentation applied to medical imaging is a longstanding computer vision problem. Deep learning has been applied successfully for some time, and the most recent advance is the emergence of a network architecture initially applied to natural language processing: transformers. At its core is the multi-headed self-attention mechanism, which compares all variables in the input sample, learning the relations between them in a fully automated method [1]. Unlike convolutional neural networks (CNN), there is no a priori imposition of any spatial relation between regions of the input images. Recently, the Swin Unet was presented [2], which applies the self-attention method together with shifted windows [3] to the encoder-decoder Unet tailored for medical image segmentation [4].

2. Materials and Methods

The used dataset is composed of 775 computed tomography (CT) scans of prostate cancer patients who had the prostate surgically removed before undergoing radiation therapy. The scans are the planning CTs used as part of their treatment. The anatomic structures segmented are the prostate surgical bed, which is the region from which the prostate had been previously removed, i.e., the radiation therapy target, and the surrounding normal tissues: bladder, rectum, and penile bulb. The scans were preprocessed so that voxel intensities match window-level and window-width of 35 and 350, using the same standard as radiologists when analyzing CTs of this anatomical region. The images were also resampled to isometric resolution of 1 mm3 voxels. A training/testing split of 700/75 samples was used. The manual segmentation performed by medical experts was used as ground-truth.

Two different models were trained: a CNN Unet, [4] and a Swin-Unet [2], both adapted for fully three-dimensional processing. The 3D volumes had size of 224 on each side. The training was performed for 150 epochs on an Nvidia DGX Station with V100 32GB GPUs.

3. Results

Using the volumetric Dice coefficient, the average across all structures and all test patients was 75.13% and 74.23% for the CNN-Unet and the Swin-Unet, respectively. However, looking at the different organs separately, the Swin-Unet outperforms in the prostate bed (68.71% vs 72.72%), and is very close in the larger organs (bladder: 96.25% vs 96.02%; rectum: 85.13% vs 84.25%), with a larger difference in the bulb (50.43% vs 43.95%). Using boundary distance metrics, the results are very close with a slight edge for the Swin-Unet: average 8.27 mm vs 7.54 mm in the 95% Hausdorff distance and 2.05 mm vs 1.90 mm in the Average Boundary Distance.

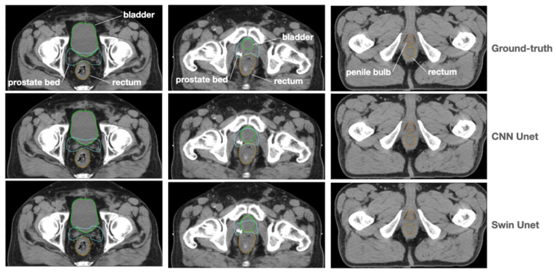

Figure 1: Examples of both neural networks' segmentations of three patients.

4. Discussion and Conclusions

A recent trend has been emerging in the Deep Learning field where there is a convergence of the techniques that were typically applied to different problems and different data types. The obtained results help corroborate this thesis, that transformer networks can be used for computer vision tasks with good outcomes. With more research and further improvements on both architectures, it is possible that a better solution lies with some mixture of both approaches.

5. References

1. Vaswani, A., et al. (2017). Attention is All you Need. Advances in neural information processing systems (NIPS), 30.
2. Cao, H., et al. (2021). Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. ArXiv, abs/2105.05537.
3. Liu, Z., et al. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 ICCV 2021, 9992-10002.
4. Ronneberger, O., et al. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015, pp. 234-241. Springer

[a] Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

[b] Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal