جامعة محمد الخامس بالرباط
Université Mohammed V de Rabat

Master de Recherche
BIO-MSCS
ENSIAS
Bioinformatique et Modélisation des Systèmes Complexes Appliquée à la Santé

ENSIAS

3D Medical images segmentation with transformers

Presented by: BELKASSMI Maryem

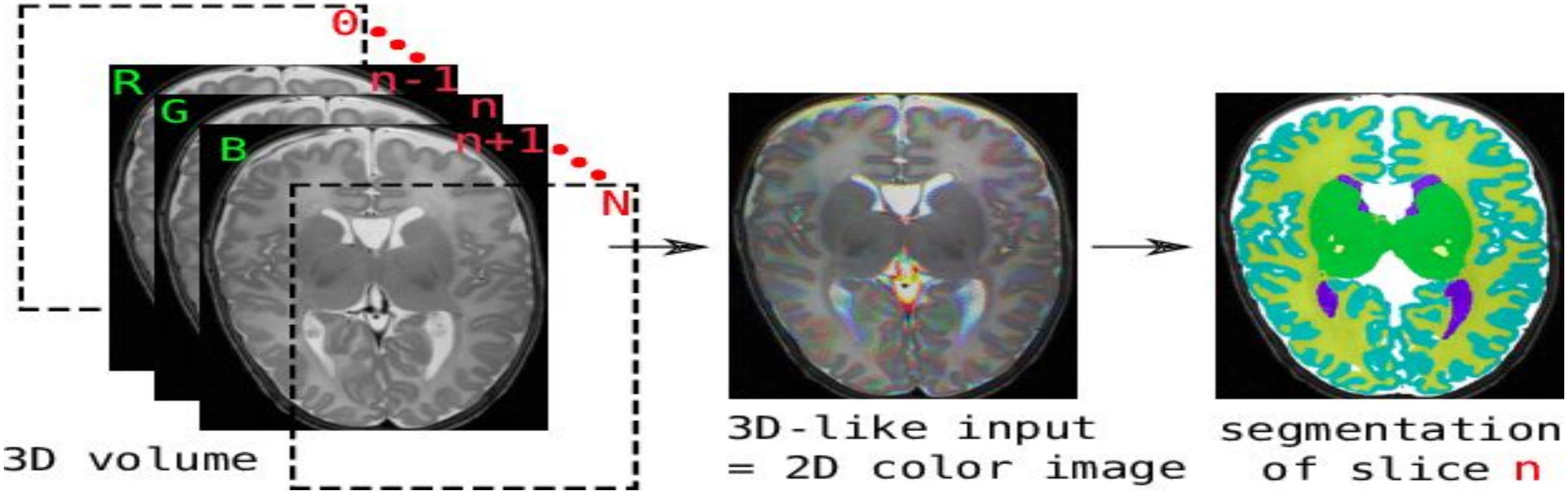
Supervised by: Pr.MAHMOUDI Abdelhak

2021-2022

Summary

- 3D medical images segmentation
- Encoder and Decoder
- Self attention models
- Transformers
- UNETR overview
- UNETR encoder-decoder architecture
- UNETR Architecture
- Loss Function
- Application on BRATS dataset
- Evaluation Metrics
- Conclusion

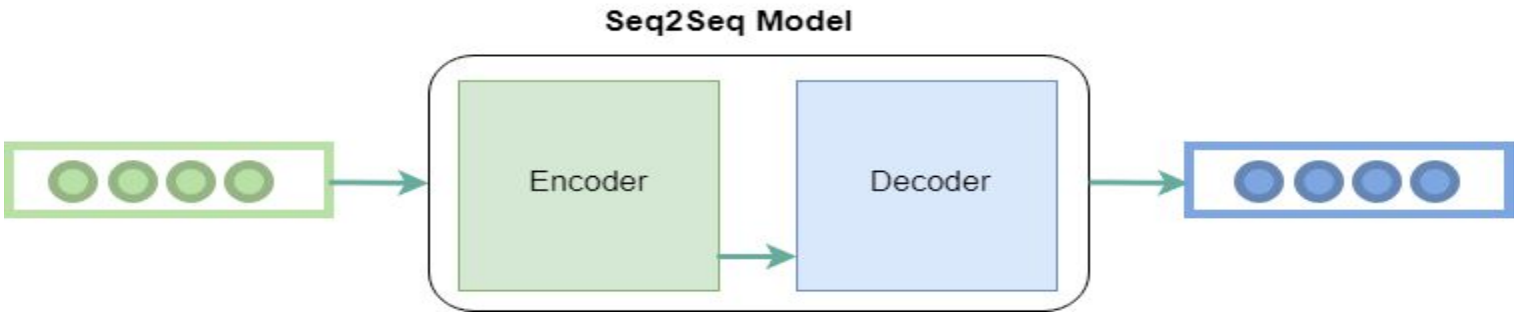
3D medical images segmentation



Encoder and Decoder

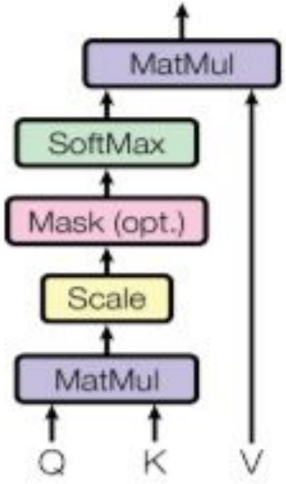
Encoder takes a variable-length sequence as the input and transforms it into a state with a fixed shape

Decoder maps the encoded state of a fixed shape to a variable-length sequence.



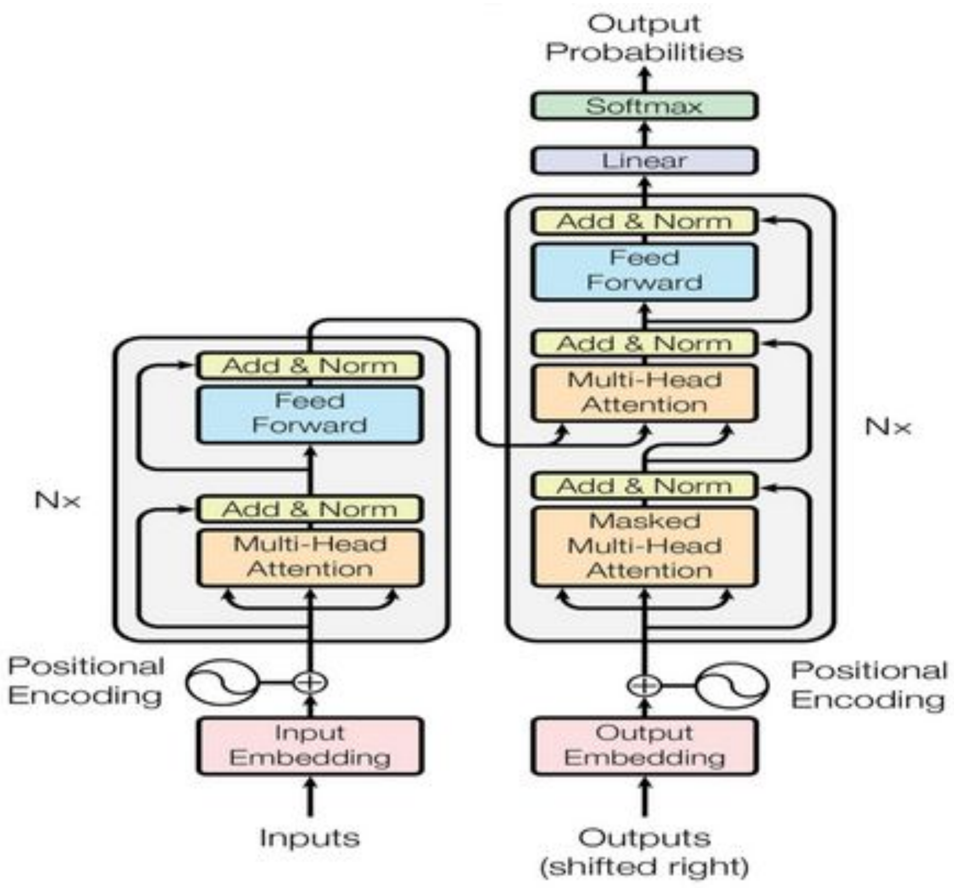
Self attention models

is an an extension of encoder and decoder mechanisme . It is proposed to overcome the limitation of the Encoder-Decoder model encoding the input sequence to one fixed-length vector from which to decode each output time step. This issue is believed to be more of a problem when decoding long sequences.



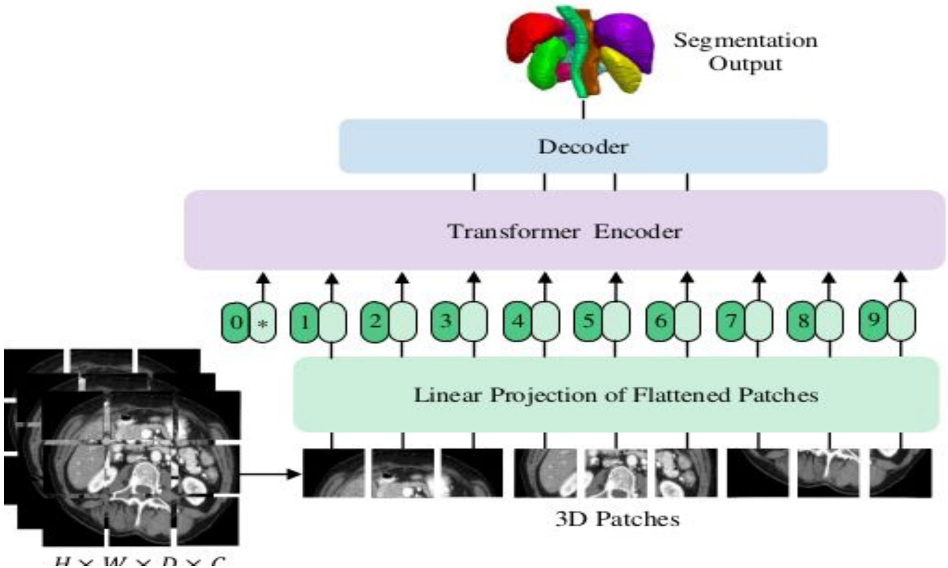
Transformers

Transformers use CNNs togethe
with **attention models.**



UNETR OVERVIEW

UNETR proposed a model that consists of a transformer encoder that directly utilizes 3D patches and is connected to a CNN-based decoder via skip connection.

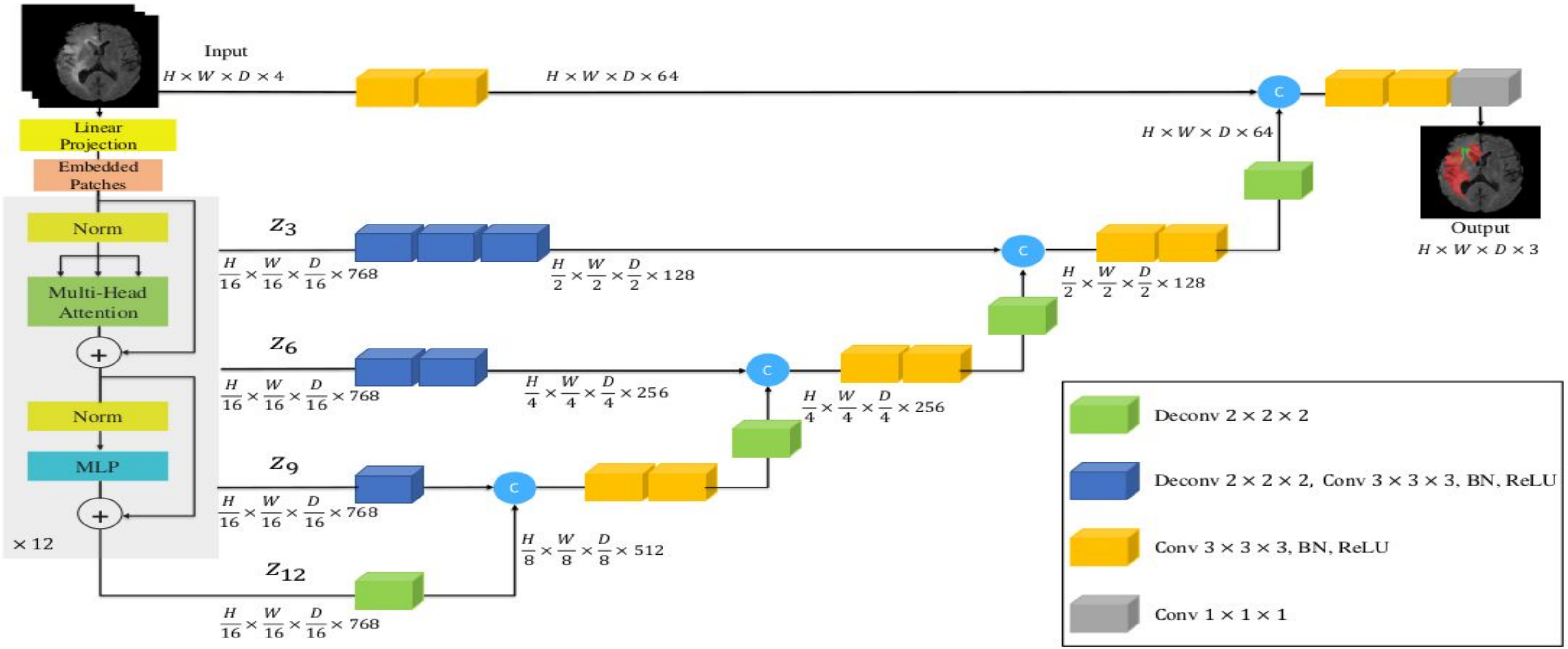


UNETR encoder-decoder architecture

- **Transformer encoder** directly utilizes the embedded 3D volumes to effectively capture long-range dependencies.

- **Skip-connected decoder** combines the extracted representations at different resolutions and predicts the segmentation output.

UNETR Architecture



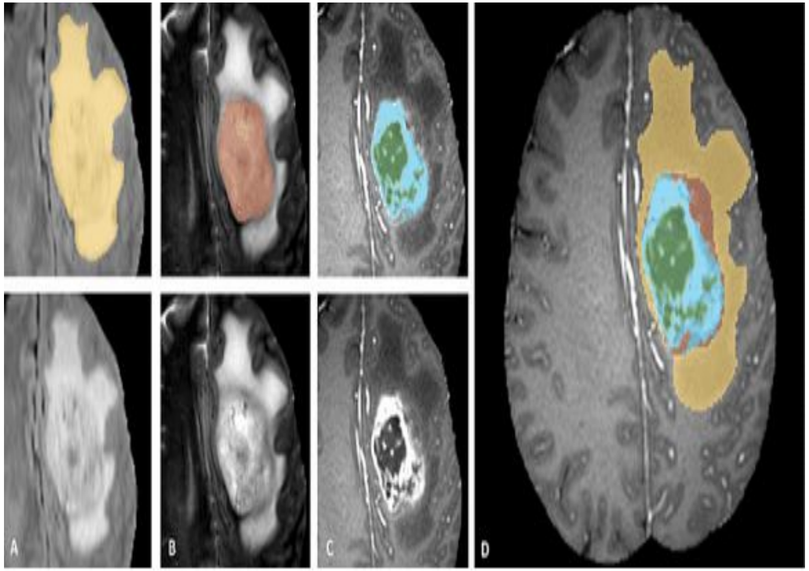
Loss Function

Loss function is a combination of soft dice loss and cross-entropy loss, and it can be computed in a voxel-wise manner according to

$$\mathcal{L}(G,Y) = 1 - \frac{2}{J}\sum_{j=1}^{J}\frac{\sum_{i=1}^{I} G_{i,j}Y_{i,j}}{\sum_{i=1}^{I} G_{i,j}^2 + \sum_{i=1}^{I} Y_{i,j}^2} - \frac{1}{I}\sum_{i=1}^{I}\sum_{j=1}^{J} G_{i,j}\log Y_{i,j}.$$

Application on BRATS dataset

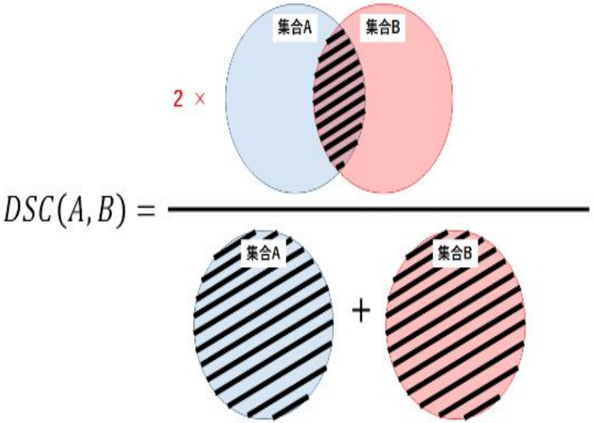
BRATS is a multi-modal large-scale 3D imaging dataset. It contains 4 3D volumes of MRI images captured under different modalities and setups. It is important to see that only the tumor is annotated.



Evaluation Metrics

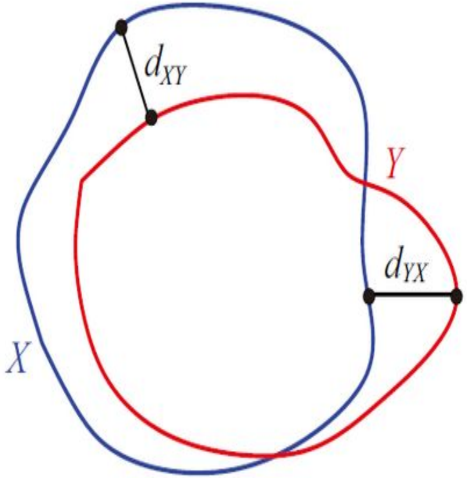
Dice Similarity Coefficient (DSC):

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \times 100\%$$



95% Hausdorff Distance (HD):

$$d_H(X,Y) = \max\{d_{XY}, d_{YX}\} = \max\left\{\max_{x \in X}\min_{y \in Y} d(x,y), \max_{y \in Y}\min_{x \in X} d(x,y)\right\}$$



Conclusion

- The paper introduces a novel transformer-based architecture, for semantic segmentation of volumetric medical images by reformulating this task as a 1D sequence-to-sequence prediction problem.
- The experiments in all datasets demonstrate superior performance of UNETR over both CNN and transformer-based segmentation models.

Ressources

https://analyticsindiamag.com/hands-on-transunet-transformers-for-medical-image-segmentation/

https://arxiv.org/abs/2103.10504

https://www.med.upenn.edu/sbia/brats2018/tasks.html

https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04

https://theaisummer.com/medical-segmentation-transformers/