Segmenting brain metastases with minimal annotations for radiotherapy application

Hugues Sibille

Semester Project

12. July 2022, Zürich

Agenda

1. Introduction
2. Materials and Methods
3. Results
4. Discussion and Conclusion

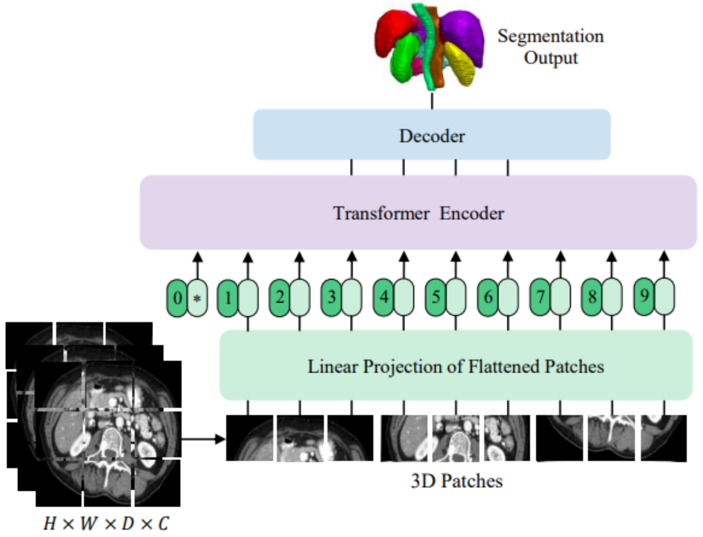
Motivation

- Medical Image Segmentation with deep learning has become a vast subject of interest since manual segmentation is a long, tedious and error-prone task.
- CNNs fixed this problem achieving close to expert‘s performance on some dataset
- Self-attention based model and specifically Vision Transformers (*Dosovitsky et al.)* [6] showed great interest for NLP tasks.
- Combining a CNN with a Vision Transformer to perform 3D image segmentation has been proposed by *A. Hatamizadeh et al* [1]
- The goal of this work is to investigate and compare how a transformer-based model can perform on our dataset.

Materials and Methods

Architecture: Vision Transformer for 3D segmentation

- Model proposed by *Hatamizadeh et al.* Named UNETR[1].
- The model uses a transformer-based encoder, which takes linearly embedded 3D patches of the input images.
- This encoder is connected to a CNN-based decoder to predict the final segmentation via skip connections.
- This model demonstrate *state-of-the-art* performance on BTCV and MSD.



Overview of UNETR
Image from [1]

Materials and Methods

Architecture: Vision Transformer for 3D segmentation[1]

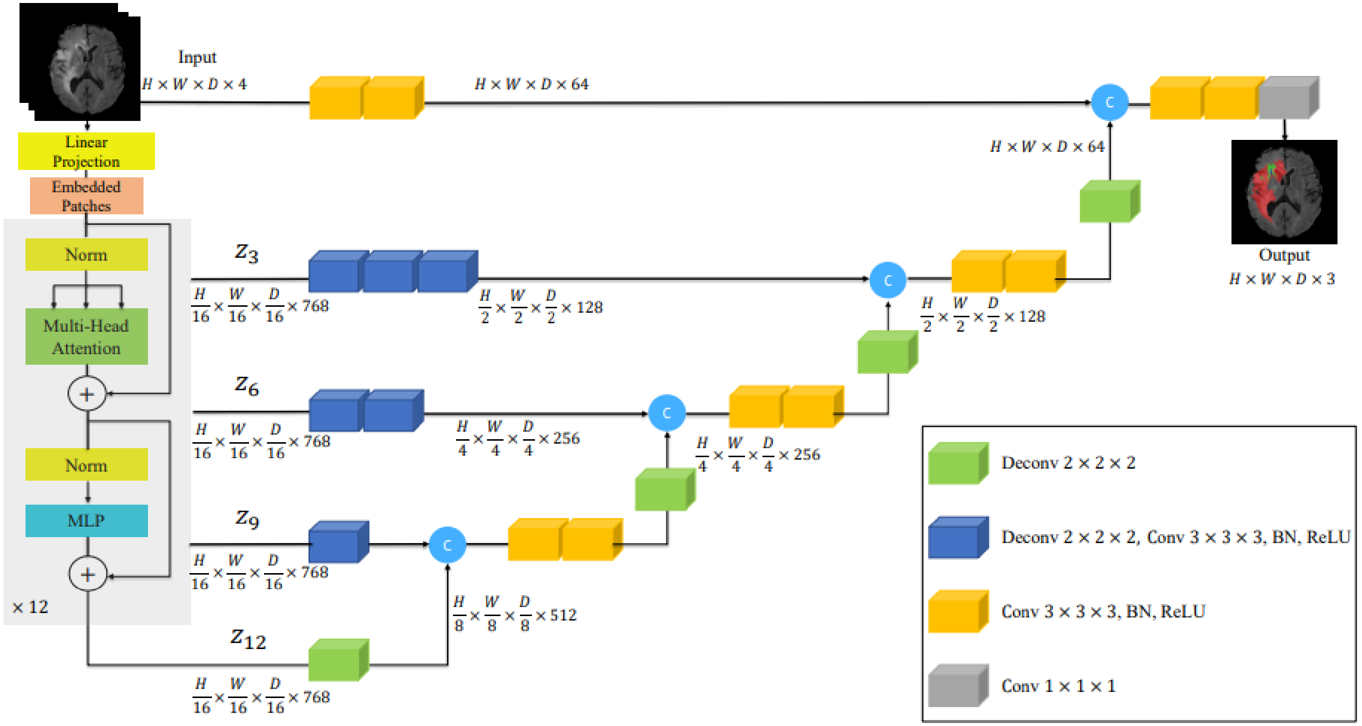


Figure 2. Overview of UNETR architecture. A 3D input volume (e.g. $C = 4$ channels for MRI images), is divided into a sequence of uniform non-overlapping patches and projected into an embedding space using a linear layer. The sequence is added with a position embedding and used as an input to a transformer model. The encoded representations of different layers in the transformer are extracted and merged with a decoder via skip connections to predict the final segmentation. Output sizes are given for patch resolution $P = 16$ and embedding size $K = 768$.

Image from [1]

Materials and Methods

Dataset

- Dataset provided by University Hospital of Zürich (USZ)
- 3D MRI of 273 patients with annotated tumors by radiologists.
- 4 types of cancers are represented in this dataset
- Each patient has a mean of 3 ± 2 metastases where the maximum is 27 metastases.
- The metastases have a mean of 22.3 ± 16.5 mm.

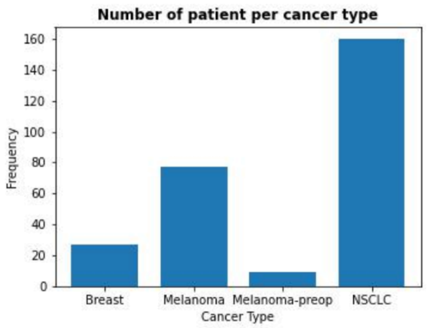
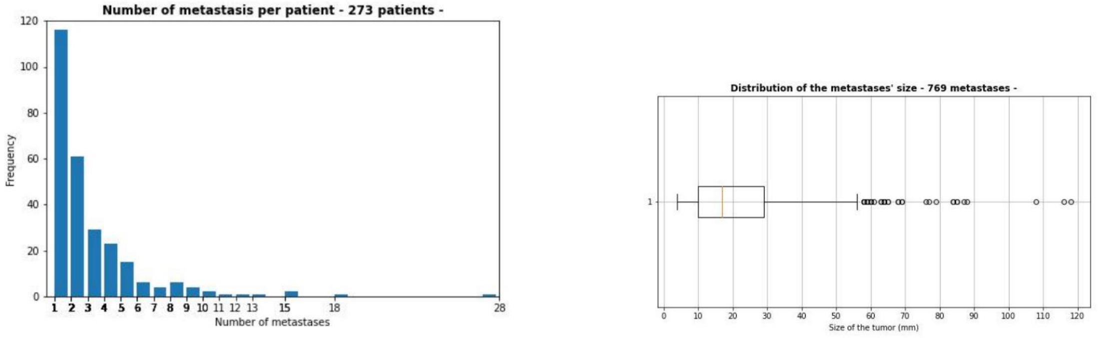


Figure 2.2: Distribution of the metastases types in the dataset.



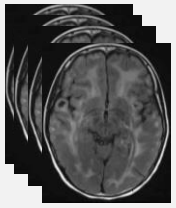
((a)) ((b))

Figure 2.3: Distribution of the number of metastases per patient across the dataset (a) and box-plot of the metastases' size in the dataset, 769 tumors has been found in the dataset (b).

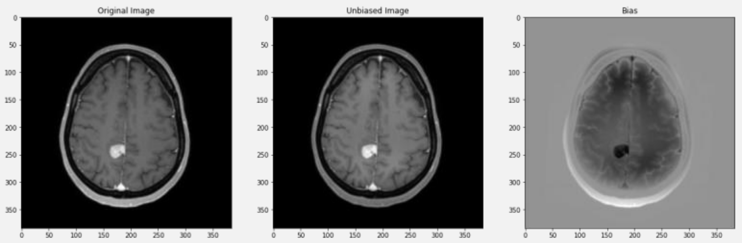
Materials and Methods

Pipeline overview

Dataset curation



Dataset of 273 3D images



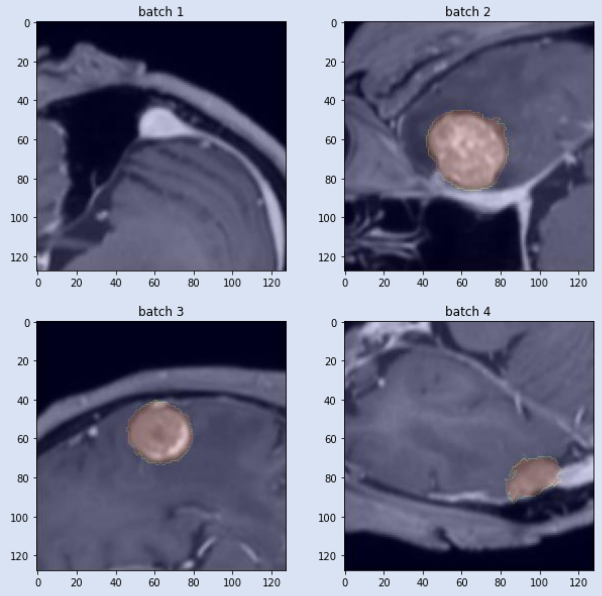
Bias Correction using N4 algorithm[5], label thresholding and outlier removal

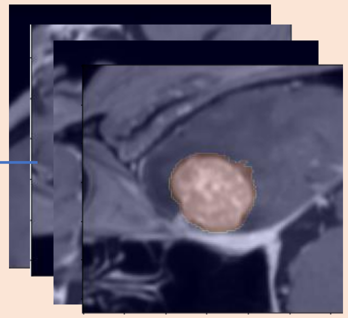
Split dataset
Training set : 217
Validation set : 54

Data preprocessing

Set of transforms from MONAI framework

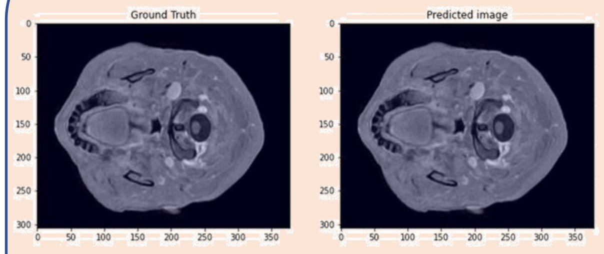
- **Normalize** and scale intensity between [0, 1]
- **Crop** the Foreground of the image
- Randomly crop 4 **[128, 128, 128] samples** from the input image
- **Data augmentation** (flip in 3 directions, rotate and scale intensity…)





Input: Batch of 4 preprocessed images of size (128, 128, 128)

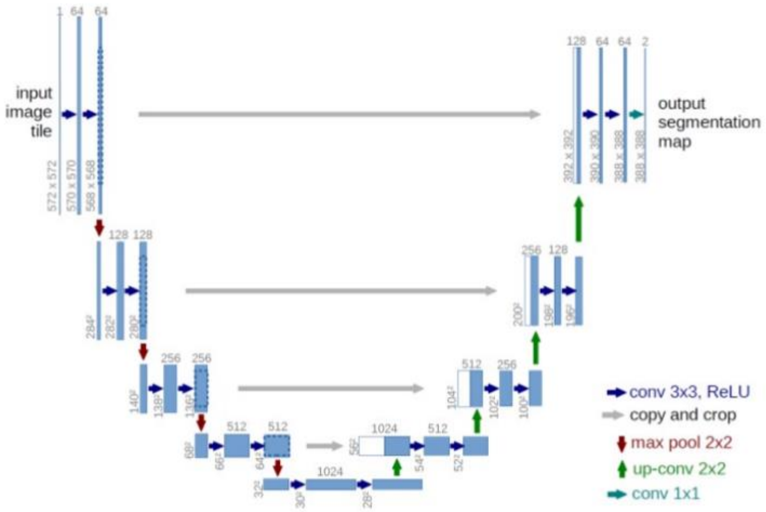
Model



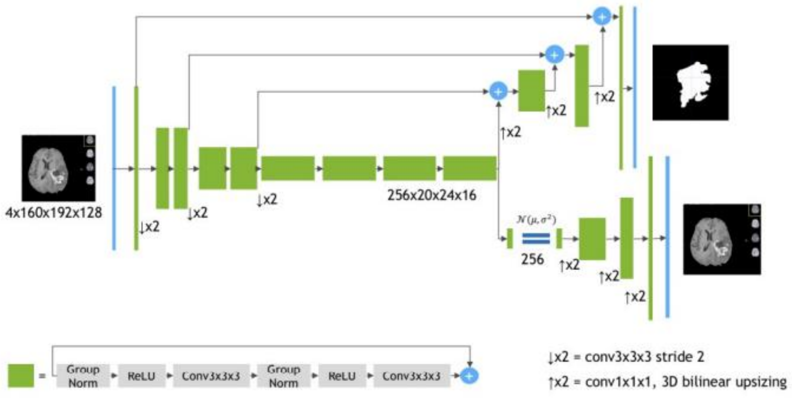
Output : Predicted Segmentation mask

Results

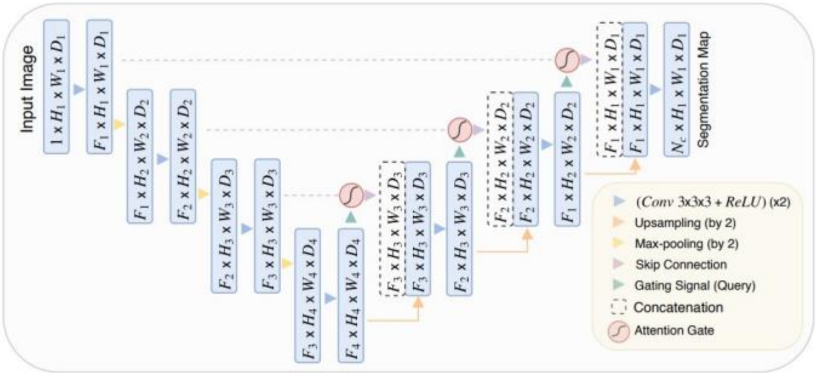
Models compared against UNETR



UNet[2]



SegResNet[3]



Attention UNet[4]

Results

Quantitative results

- Dice Score Coefficient (DSC), 95th Hausdorff distance, Precision and Recall are used to validate the models
- ViT-based model (UNETR) was outperformed by UNet-based models.
- Attention Unet showed best results among the four.
- 5-fold cross validation was performed and once again Attention-UNet showed the best metrics on all splits.

	UNETR	UNet	SegResNet	Attention-UNet
DSC	0.596±0.350	0.723±0.286	0.711±0.284	**0.756±0.278**
HD95 (mm)	7.090±7.066	5.507±6.705	8.348±6.762	**5.127±6.155**
Precision	0.739±0.324	0.773±0.282	0.751±0.276	**0.787±0.283**
Recall	0.598±0.344	0.745±0.293	0.762±0.279	**0.794±0.252**

Table 3.1: Comparison of different models on the segmentation task. Those results were obtained with the validation set used during training. The best results are highlited in bold.

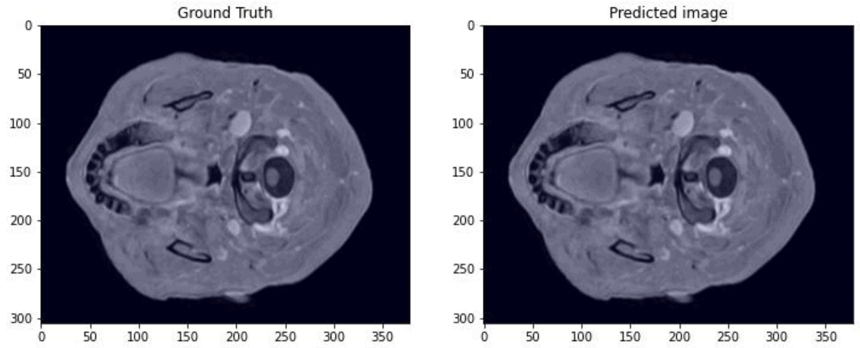
	UNETR		UNet		SegResNet		Attention UNet	
	DSC	HD95	DSC	HD95	DSC	HD95	DSC	HD95
Split 0	0.655	8.698	0.719	5.209	0.699	9.671	**0.764**	**4.891**
Split 1	0.663	10.161	0.703	5.041	0.642	10.211	**0.778**	**4.803**
Split 2	0.637	9.219	0.719	5.471	0.693	9.283	**0.766**	**4.661**
Split 3	0.687	9.385	0.731	4.116	0.713	8.617	**0.781**	**4.195**
Split 4	0.732	8.991	0.774	4.062	0.751	8.579	**0.813**	**4.851**
Average	0.675	9.291	0.729	4.780	0.700	9.272	**0.780**	4.680

Table 3.2: 5-fold cross validation test of different models. The best results are highlited in bold.

Results

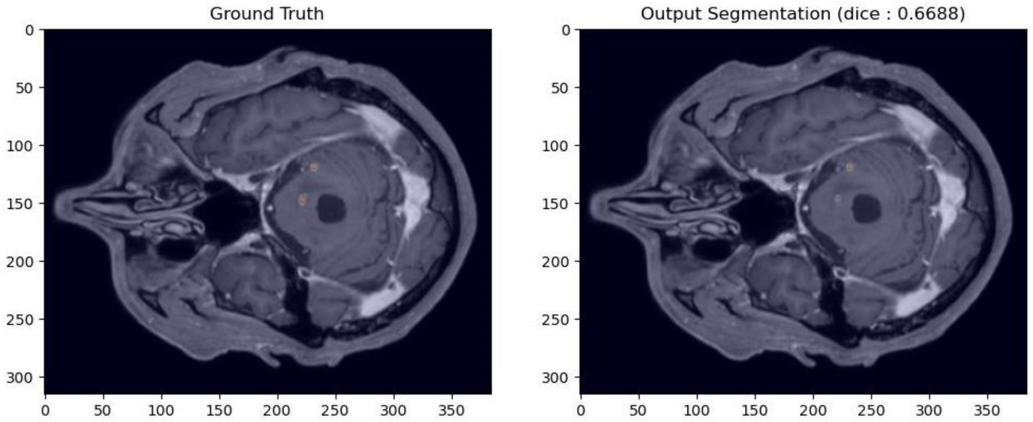
Qualitative results : UNETR *(A. Hatamazideh et al.)*[1]

- Trained for 1400 epochs (~120h)
- Showed promising results
- Best *Val Mean Dice : 0.5755*

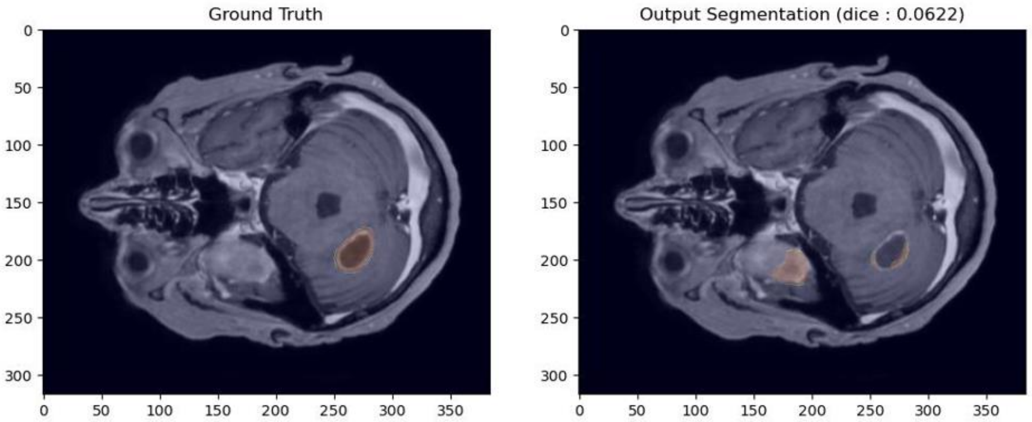


Good prediction by UNETR : Dice score **0.9416**

False Negative



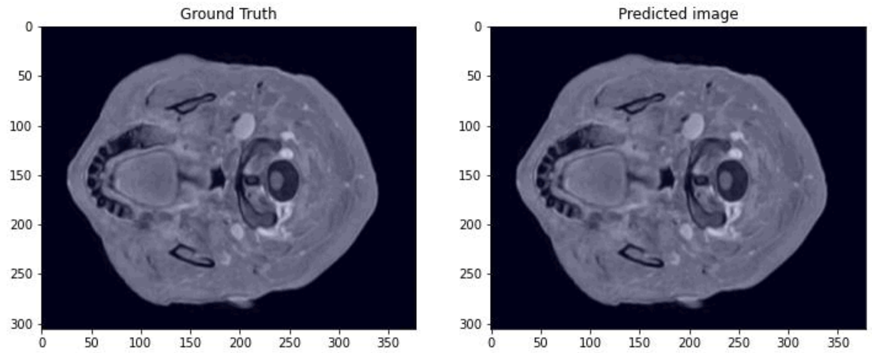
False Positive



Results

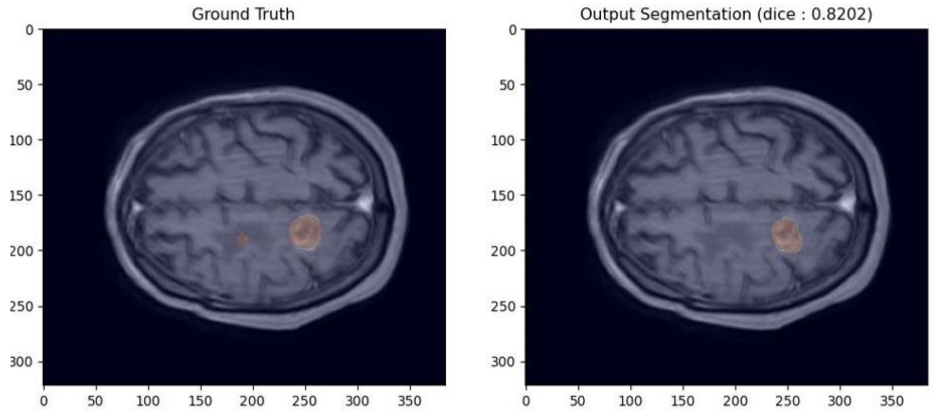
Qualitative results : **U-Net** *(O. Ronneberger et al.)* [2]

- Trained for 350 epochs (~30h)
- Best *Val Mean Dice : 0.7133*
- Perform well on this dataset, small tumors are harder to segment and lower part of the brain can be confusing for the network.

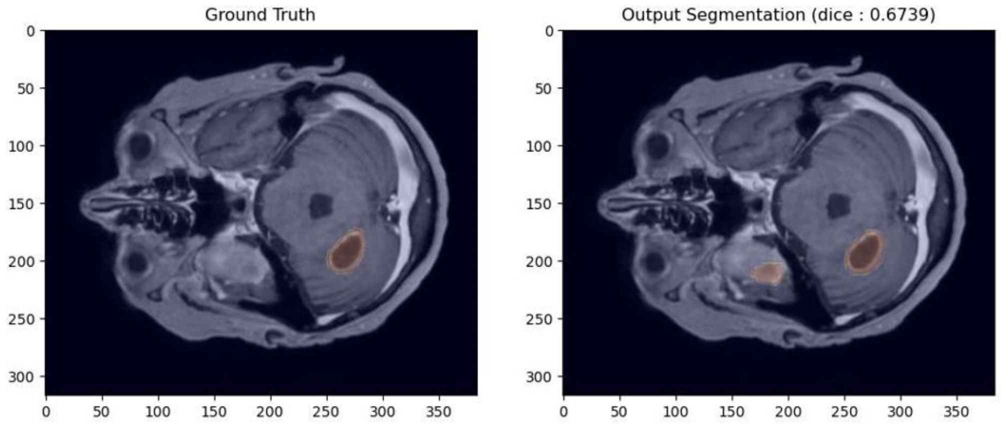


Good prediction by UNet : Dice score **0.9538**

False Negative



False Positive



Results

Qualitative results

- Overall good segmentation by each network even if UNETR seems quite below the three others.
- Segmentation is harder on the lower part of the brain, and small structures are difficult to segment for all networks.

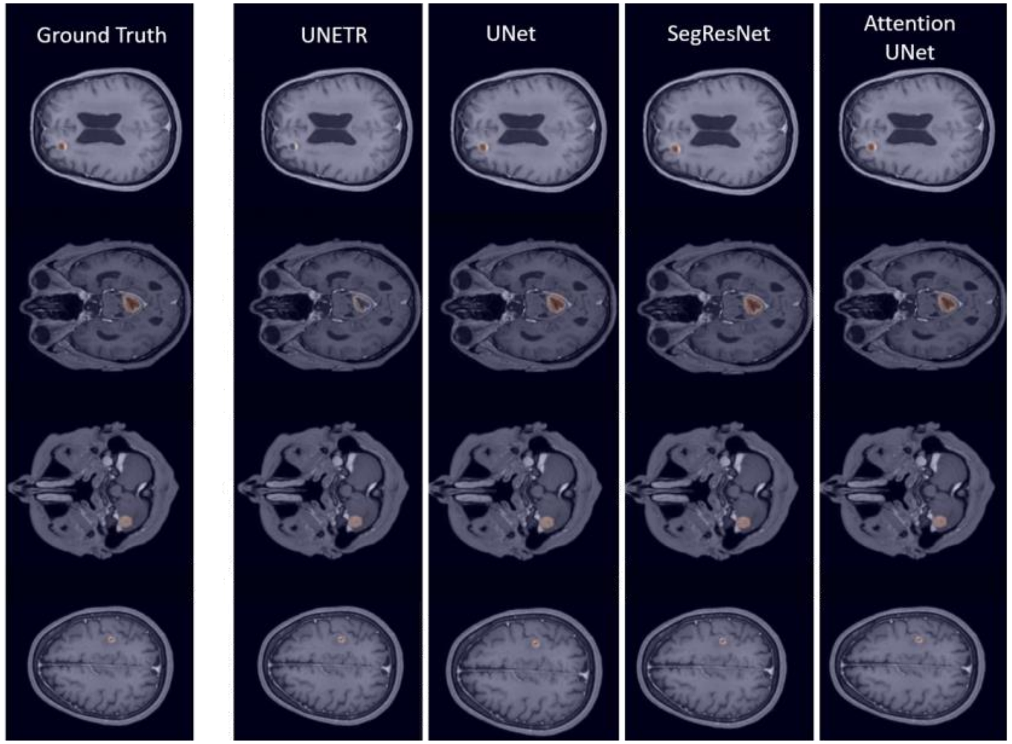


Figure 3.1: Visualisation of some interesting slices from different patient, where there is at least one tumor segmented. The segmentation mask is in brown here.

Discussion and Conclusion

- We compared three models against the model of interest, which was Vision Transformers.
- The pipeline presented allowed flexibility for the choice of the model, the type of structures we want to segment, and parameter tuning.
- Using Transformers as an encoder can segment medical images and show reasonably good results compared to related literature.
- However, the three others outperformed UNETR, so this model doesn’t seem to be the best model for this dataset.
- Unet-based models and mainly the Attention-Unet showed the best results along the project and therefore can be interesting to build upon their recommendation

References

[1] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation, 2021.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[3] Andriy Myronenko. 3d MRI brain tumor segmentation using autoencoder regularization. CoRR, abs/1810.11654, 2018

[4] Ozan Oktay, Jo Schlemper, Loïc Le Folgoc, Matthew C. H. Lee, Mattias P. Heinrich, Kazunari Misawa, Kensaku Mori, Steven G. McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. CoRR, abs/1804.03999, 2018.

[5] Nicholas J et al. Tustison. N4itk: improved n3 bias correction. IEEE transactions on medical imaging vol. 29, pages 1310–20, 2010

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, 2020.