

# ATTENTION (AND DATA AUGMENTATION) IS ALL YOU NEED TO IMPROVE TRANSFORMER-BASED 3D BRAIN SEGMENTATION

*Siddhant Sanghi, Shashwat Singh and Matthew Spotnitz*

Columbia University

## ABSTRACT

Differentiating among brain tissues in medical imaging is an active area of research. Previously, convolutional neural network (CNN) architectures like U-Net have been implemented for tissue segmentation tasks but with limited success. Subsequently, transformers have been used to improve tissue segmentation performance, in models like the Transformer-based Automated Brain-tissue Segmentation (TABS) model [1]. We wish to increase TABS test accuracy and performance by implementing MRI-artifact inspired online data augmentation and self-attention through Squeeze-Excitation (SE) blocks.

*Index Terms— MRI, Brain Tissue Segmentation, Transformer, Data Augmentation, Squeeze-Excitation*

## 1. INTRODUCTION

Differentiating among different types of brain tissue on magnetic resonance images (MRIs) is an essential part of interpreting those images. Specifically, distinguishing among gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) is necessary [1]. To be able to survey images from healthy subjects to quantify morphology and margins of these tissues in MRI scans through segmentation is essential in order to provide health benchmarks with the goal of being able to identify abnormalities in such scans.

Abnormalities in the volume of different kinds of tissue are an important component of a range of neurological or psychiatric diseases. Specifically, Schizophrenia, Alzheimer’s Disease, Huntington’s Disease, and Bipolar Disorder are associated with structural abnormalities in the brain [1]. Furthermore, tissue segmentation could help classify normal and abnormal tissue. That differentiation could help physicians prioritize which segments of a brain MRI warrant the most focus. Overall, effective tissue segmentation could improve the accuracy and consistency

of clinical diagnoses. Different approaches to tissue segmentation have been described in the prior literature, but with limited success.

## 2. REVIEW OF PREVIOUS WORK

Early research efforts focused on intensity thresholding, population-based atlases, statistical methods and standard machine learning algorithms. However, the efficacy of those approaches was limited by the quality of the inputs or *a priori* information [1].

Convolutional neural networks (CNNs) have shown superior performance to standard machine learning algorithms for brain tissue segmentation, due in part to their feature-encoding capabilities [1]. Although the CNNs have performed well on training data sets, their generalizability to new data sets has been limited. However, the use of transformers along with CNNs, such as Transformer-based Automated Brain tissue Segmentation (TABS), has shown promise as a brain segmentation technique [1].

## 3. DATASET AND CODE AVAILABILITY

We used the SALD [2] dataset. This dataset contains cross-sectional brain MRI images. We have modified the existing codebase of TABS [3]. This is a CNN-Transformer based brain tissue segmentation architecture which relies on a Res-UNet backbone. The code is available in a Github repository of the authors of this paper [4].

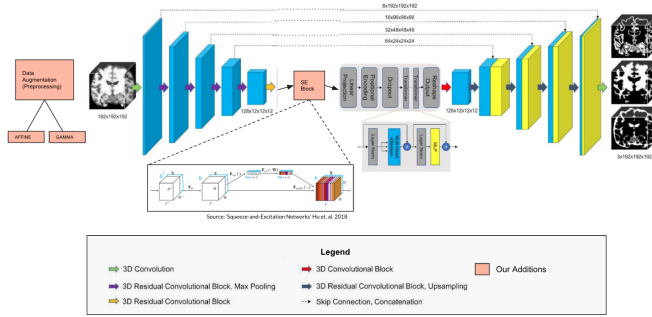
In total we had 283 3D images which were further divided into 5 folds (2 folds with 56 images and 3 folds of 57 images). Additionally, we had 849 mask images representing CSF, WM and GM.

## 4. METHODS

As we only had 170 3D images as training data and as it is known that these complex neural networks easily tend to

overfit for small training data sample size, we wanted to implement data augmentation for 3D brain MRI images. For data augmentation we particularly selected two kinds of online data augmentation methods mimicking realistic artifacts in MRI images from the torchio library. We chose only two methods due to Kaggle memory constraints.

Additionally, we added Squeeze - Excitation blocks after the final residual convolutional block. The complete modified TABS architecture is shown in *Fig. 1*.

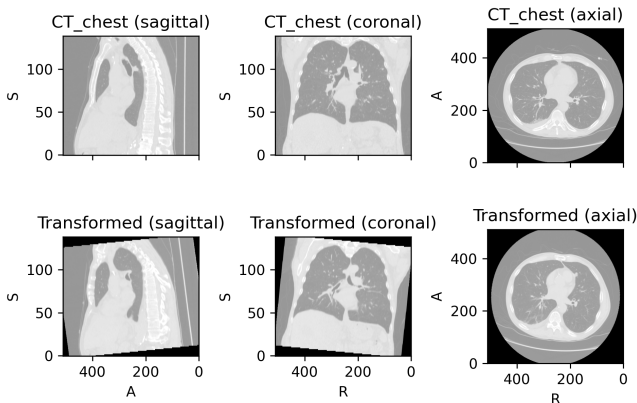


**Fig. 1.** Modified TABS Model Architecture

#### 4.1. Data Augmentation

##### 4.1.1. Random Affine Transformation

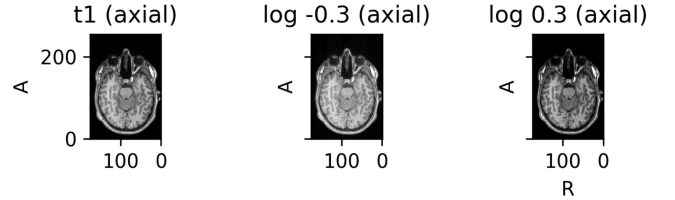
Random affine transformations aim to interpolate the image intensity values. Such transformations are used for many different purposes including simulating the motion blur effect during MRI scans. They can be used to scale original images in addition to rotating them by custom angles. Images can also be translated in 3D space along with the option of providing padding for the resultant image. An example of this transformation from the official torchio website [5] is shown below in *Fig. 2*.



**Fig. 2.** Representation of Random Affine Transformation

##### 4.1.2. Random Gamma Transformation

Random gamma transformations are used to change the contrast of the image by changing the values of gamma randomly. They usually raise the image value by the power of gamma. This augmentation method mimics contrast changes due to Cross-excitation MRI Artifacts [6]. An example of this transformation from the official torchio documentation [5] is shown in *Fig. 3*.

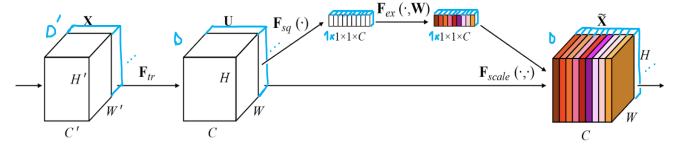


**Fig. 3.** Representation of Random Gamma Transformation

#### 4.2. Feature Recalibration through Self-Attention

##### 4.2.1. 3D Squeeze-Excitation Blocks

We implemented a 3D extension to the 2D Squeeze-Excitation (SE) Block as seen in *Fig. 4*, where we added a Depth (D) dimension to the HxWxC 2D feature.



**Fig. 4.** 3D Squeeze-Excitation Block (extension to 2D version [7] annotated in blue)

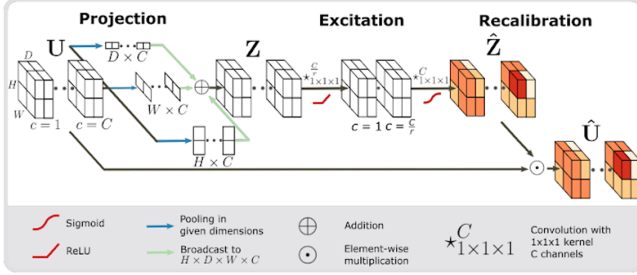
We chose SE as a method to improve TABS by allowing dynamic channel-wise feature recalibration through self-attention after residual convolution blocks.

Due to Kaggle memory limitations, we were able to add only a single SE block after the Encoder at the end of convolutional layer 5 to the TABS architecture, as seen in *Fig. 1*. The 'Squeezing' was done through an average pooling for 3D. To create the weighting for 'Excitation' there was a small side network comprising a Linear Layer followed by ReLU, followed by reduction of output channel complexity by ratio of 8, followed by another Linear Layer and sigmoid smoothing. The 'Excitation' was weighting each feature of the convolution block based on the side network.

##### 4.2.2. Project-and-Excite Blocks

According to a recent paper published by Rickmann et. al. [8], the 3D-extension of the 2D SE block may not be optimal for 3D image segmentation tasks, since it was designed for 2D.

Hence they suggested the creation of Project-and-Excite (PE) block as depicted in Fig. 5, combining spatial and channel information for recalibration, through projection onto  $D \times C$ ,  $W \times C$  and  $H \times C$  dimensions followed by broadcasting to the original  $H \times D \times W \times C$ , with specific application in the realm of 3D image segmentation.



**Fig. 5.** Project-and-Excite Block [8]

The PE block was proposed to retain more of the valuable spatial information within the 'Project' operation, following which the 'Excitation' would learn inter-dependencies between the projections across the different channels, learning not just channel-wise information as in SE, but spatial information too.

Unfortunately, the PE block did not admit as good a performance as 3D-extension of the 2D SE block did, in 25-epoch trials using the two blocks individually in the same location in the network. We believe that the poorer PE block performance is due to a 7 times larger training dataset in this paper, than in Rickmann et al.'s paper.

## 5. RESULTS AND DISCUSSION

We measured the performance on multiple metrics. These metrics can be divided into Probability-based Metrics as shown in Tables A, C and E and Map-based Metrics as shown in Tables B, D and F. These results are shown over 120 epochs for Baseline TABS without any modifications (Tables A and B), for our modified TABS model with data augmentation only (Tables C and D) and our modified architecture with both data augmentation and SE block (Tables E and F).

We chose to add our modifications to the code stepwise to make sure we were improving accuracy at each stage. We made comparisons of our modifications to the Baseline for 120 epochs (Tables A and B).

**Table A.** Probability-based Metrics for Baseline TABS Model for 120 epochs

	Gray Matter	White Matter	CSF

Pearson	0.612	0.633	0.624
Spearman	0.498	0.611	0.602
pSNR	10.982	9.899	10.692
MSE	0.122	0.234	0.138

**Table B.** Map-based Metrics for Baseline TABS Model for 120 epochs

	Gray Matter	White Matter	CSF
DICE	0.684	0.707	0.699
HD	25.547	22.213	28.653
Jaccard	0.519	0.364	0.957

**Table C.** Probability-based Metrics for our Modified TABS Architecture with only Data Augmentation for 120 Epochs

	Gray Matter	White Matter	CSF
Pearson	0.686	0.723	0.817
Spearman	0.616	0.715	0.738
pSNR	8.812	7.486	8.554
MSE	0.074	0.102	0.077

**Table D.** Map-based Metrics with only Data Augmentation for 120 Epochs

	Gray Matter	White Matter	CSF
DICE	0.774	0.758	0.758
HD	20.850	18.094	25.459
Jaccard	0.711	0.682	0.692

We observed that our model with data augmentation only (Tables C and D) had higher performance metrics overall than baseline, to indicate that data augmentation is a useful

strategy in improving 3D image segmentation performance of TABS.

**Table E.** Probability-based Metrics with Data Augmentation and the Squeeze-Excitation block for 120 epochs

	Gray Matter	White Matter	CSF
Pearson	0.705	0.709	0.827
Spearman	0.621	0.799	0.726
pSNR	9.879	8.461	9.605
MSE	0.072	0.084	0.067

**Table F.** Map-based Metrics with Data Augmentation and the Squeeze-Excitation block for 120 epochs

	Gray Matter	White Matter	CSF
DICE	0.751	0.734	0.791
HD	20.814	18.030	25.881
Jaccard	0.495	0.465	0.521

Then, we added the SE block as shown in *Fig. 1* and trained for 120 epochs with both data augmentation and the SE block (*Tables E and F*).

We observed that we achieved better performance metrics on segmenting CSF than segmenting White Matter and Grey Matter, than the baseline (*Tables A and B*) and baseline with only data augmentation (*Tables C and D*). This may be happening because the model only ran with one single SE block after encoder convolutional layer 5, which might be having better feature embeddings for segmenting the CSF better than the White Matter and Grey Matter, indicating possibly that the layer before it focused more on segmenting CSF than the other tissues.

Finally, we were able to see improved accuracy from baseline with just 120 epochs of training, while originally, it took 350 epochs to reach higher state-of-the-art results. Thus, it is our belief that with a larger number of epochs our modified TABS architecture can outperform the state-of-the-art TABS performance.

## 6. CONCLUSION

We were majorly constrained with memory and time-limit constraints of Kaggle, limiting runs to no more than 120 epochs, and with just two data augmentation methods and a single SE block. But we believe that our model can outperform the original TABS model if it is run for more epochs on a strong GPU (outside of Kaggle), with SE blocks following each convolutional block and more data augmentation methods to mimic MRI artifacts.

## 7. ACKNOWLEDGEMENTS

We are grateful for the assistance from Dr. Jia Guo, Vishwanatha M. Rao, Soroush Arabshahi, Xuzhe Zhang, and the creators of the SALD dataset [2].

## 8. REFERENCES

- [1] Rao VM, Wan Z, Ma DJ, Lee P-Y, Tain Y, Laine AF, Guo J. IMPROVING ACROSS-DATASET BRAIN TISSUE SEGMENTATION USING TRANSFORMER. (2022). *Arxiv*.
- [2] Wei, D., Zhuang, K., Ai, L. et al. STRUCTURAL AND FUNCTIONAL BRAIN SCANS FROM THE CROSS-SECTIONAL SOUTHWEST UNIVERSITY ADULT LIFESPAN DATASET. (2018). *Sci Data* 5, 180134.
- [3] <https://github.com/raovish6/tabs>
- [4] [https://github.com/s10singh97/BME\\_Final\\_Project](https://github.com/s10singh97/BME_Final_Project)
- [5] “Augmentation,” TorchIO. [Online]. Available: <https://torchio.readthedocs.io/transforms/augmentation.html>.
- [6] M. Saber, “Cross-excitation artifact (MRI): Radiology reference article,” Radiopaedia Blog RSS, 15-Jun-2021. [Online]. Available: <https://radiopaedia.org/articles/cross-excitation-artifact-mri-1>.
- [7] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, Enhua Wu. SQUEEZE-AND-EXCITATION NETWORKS. (2017). *Arxiv*.
- [8] Anne-Marie Rickmann, Abhijit Guha Roy, Ignacio Sarasua, Nassir Navab, Christian Wachinger. PROJECT & EXCITE MODULES FOR SEGMENTATION OF VOLUMETRIC MEDICAL SCAN. (2019). *Arxiv*.