

Article

Deep 3D Neural Network for Brain Structures Segmentation Using Self-Attention Modules in MRI Images

Camilo Laiton-Bonadiez ^{1,*}, German Sanchez-Torres ² and John Branch-Bedoya ¹¹ Facultad de Minas, Universidad Nacional de Colombia, Medellín 050041, Colombia; jwbranch@unal.edu.co² Facultad de Ingeniería, Universidad del Magdalena, Santa Marta 470001, Colombia; gsanchez@unimagdalena.edu.co

* Correspondence: claiton@unal.edu.co

Abstract: In recent years, the use of deep learning-based models for developing advanced healthcare systems has been growing due to the results they can achieve. However, the majority of the proposed deep learning-models largely use convolutional and pooling operations, causing a loss in valuable data and focusing on local information. In this paper, we propose a deep learning-based approach that uses global and local features which are of importance in the medical image segmentation process. In order to train the architecture, we used extracted three-dimensional (3D) blocks from the full magnetic resonance image resolution, which were sent through a set of successive convolutional neural network (CNN) layers free of pooling operations to extract local information. Later, we sent the resulting feature maps to successive layers of self-attention modules to obtain the global context, whose output was later dispatched to the decoder pipeline composed mostly of upsampling layers. The model was trained using the Mindboggle-101 dataset. The experimental results showed that the self-attention modules allow segmentation with a higher Mean Dice Score of 0.90 ± 0.036 compared with other UNet-based approaches. The average segmentation time was approximately 0.038 s per brain structure. The proposed model allows tackling the brain structure segmentation task properly. Exploiting the global context that the self-attention modules incorporate allows for more precise and faster segmentation. We segmented 37 brain structures and, to the best of our knowledge, it is the largest number of structures under a 3D approach using attention mechanisms.

Keywords: medical image segmentation; deep learning; transformers; convolutional neural networks; brain structures



Citation: Laiton-Bonadiez, C.; Sanchez-Torres, G.; Branch-Bedoya, J. Deep 3D Neural Network for Brain Structures Segmentation Using Self-Attention Modules in MRI Images. *Sensors* **2022**, *22*, 2559. <https://doi.org/10.3390/s22072559>

Academic Editor:

Alessandro Bevilacqua

Received: 7 February 2022

Accepted: 21 March 2022

Published: 27 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The scientific community has developed tools that allow for obtaining brain information so doctors can study the human brain. Among these tools we find brain imaging, which includes methods such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and ultrasound (US), among others. However, not all methods produce quality images when applied to the brain because high contrast images are required to study the human brain. Therefore, highly sensitive methods such as magnetic resonance imaging or positron emission tomography must be used [1].

Although PET scans are capable of obtaining good quality images, they are not the preferred choice for specialists as they have several disadvantages. PET scans cannot reveal structural information at the microscopic and macroscopic levels in the white and gray matter of the brain; cannot detect changes in brain activation, and pose health risks due to the required radiation [2]. For these reasons, the MRI method has been widely used in the brain for medical studies and scientific research [3,4]. There are several types of MRI sequences that are capable of improving contrast and brightness in certain types of tissues. T1-weighted, T2-weighted, Fluid Attenuated Inversion Recovery (Flair), and Diffusion Weighted Imaging (DWI) are among the most common MRI sequences [5].

Traditionally, interpretations of medical images have been made by human experts. However, the existence of variations in the criteria among various human specialists is a limitation in relation to the generation of an efficient and precise diagnosis [6]. This weakness has been addressed through the development of computer aided diagnostic systems using computer vision for analysis of medical Images [7]. In this field, traditional algorithms were initially applied for the segmentation of brain anatomical structures, such as thresholding techniques [8], growth of regions [9], machine learning algorithms for classification [10], or grouping [11], among others. Based on the growth in computational capacity and the amount of data available, it is possible to use more robust and complex modern algorithms achieving better results in medical segmentation tasks [12].

In general, multiple methods for the segmentation of brain magnetic resonance images have been proposed. These methods can be grouped as follows: manual methods, spatial dimensionality-based methods (2D and 3D), pixel/voxel intensity-based methods, atlas-based methods, surface-based methods, and methods based on deep learning techniques [13].

The manual segmentation of magnetic resonance images is based on the use of highly trained personnel to obtain the different types of brain tissues. In order to perform segmentation, experts commonly use manual delineation tools that allow them to delineate different regions of the brain. Some examples of these tools are FreeSurfer [14], BrainSuite [15], FSL [16], ITK-SNAP [17], 3D Slicer [18], SPM [10], and Horos [19], among others.

In terms of the spatial dimensionality methods, these are subdivided into 3D and 2D approaches. Three-dimensional-based segmentation approaches seem to be the natural way to approach the problem because it allows to exploit the three-dimensional nature of MRI by considering each voxel and its relationship to neighbors at different acquisition planes (sagittal, coronal, and axial). However, the 3D approach still has limitations, mainly related to the high computational cost in computers with limited memory, the increase in the complexity of the models and the number of parameters, making the learning process slower [20–22]. Therefore, researchers usually use the 2D representation of a brain MRI to avoid memory restraints and computational limitations of the 3D representation.

Intensity-based methods attempt to find a threshold value that separates the different tissue categories. These methods include techniques of thresholding, growth of regions, classification, and grouping [8]. In [23,24], authors presented work on pixel intensity using thresholding techniques to segment brain tumors.

Works that use region growth techniques, use the similar characteristics of pixels found together to perform the separation of a common region [25]. Region growth techniques have been applied for the segmentation of brain tumors [9], organs [26], cerebral vessels [27], and lesions in both the brain and the breast, applying other techniques such as morphological filters [28] or with quantitative values such as the measurement of the roughness of the tumor border [29]. Within this group, we also find classification and grouping techniques that make use of labeled and unlabeled data for their operations. In fact, multiple brain segmentation tools that are widely used in the scientific community, such as FreeSurfer [14], 3D Slicer [18], and SPM [10] make use of Bayesian Classifiers. The k-means algorithm is the most used in clustering because it is simple to implement, relatively fast, and produces good results. Some examples of work for tissue or tumor segmentation are presented in [11] and in [30], where the k-means algorithm is combined with a vector support machine and a Fuzzy C-means algorithm with thresholding techniques, respectively.

Atlas-based methods are those that make use of brain anatomical information from a specific population for image segmentation. Likewise, in [31–34], it is shown that this method is not limited to the use of a specific atlas; multiple atlases and techniques such as tag fusion can also be used to improve the delimitation of brain regions.

Several works have also been proposed that make use of deformable models. These techniques are part of the group of surface-based methods where the main objective is the delimitation of regions with similar characteristics through the use of elastic curves [35].

Similar to the approaches mentioned above, surface-based methods have been used for the segmentation of brain regions [36] and tumors [37,38].

Recently, deep learning has become an area of interest in the scientific community due to the important results that have been achieved in multiple disciplines [39–43]. The strength of convolutional networks is that they automatically identify relevant characteristics without any explicit human supervision [44]. In addition, compared with their predecessors, fully connected neural networks, convolutional networks significantly reduce the number of trainable network parameters, facilitating computation and making it possible to build large-scale architectures. Likewise, they efficiently link the output of the network with the extracted characteristics by jointly training the classification layers and the characteristic extraction layers [45]. Specifically, in the problem of brain magnetic resonance imaging segmentation, deep learning has achieved better results than previously exposed methods [12]. Within the deep learning branch, there are multiple algorithms based on neural networks that have been developed with specific objectives, such as autoencoders [46], Boltzmann machines [47], recurrent neural networks [48], convolutional neural networks [49], and Transformers [50], among others. Convolutional neural networks are precisely the algorithms most widely used by researchers to perform image segmentation and classification tasks, given that they have achieved the best results to date.

Convolutional neural networks are a type of neural network that was created by LeCun but was inspired by Fukushima's work on the neocognitron for the recognition of handwritten Japanese characters [51]. In the study of brain magnetic resonance imaging using neural network architectures, it is common to see convolutional neural networks as the basis of the architectures. In fact, in [52], the authors presented a solution for brain tissue segmentation on MRI images taken in infants approximately six to eight months of age using CNNs in the deep learning architecture.

Similarly, the authors of [53] were able to use CNNs for the segmentation of subcortical brain structures using the datasets of Internet Brain Segmentation Repository (IBSR) and LPBA40 [15].

Some of the most important deep learning solutions have been proposed using a 2D representation allowing researchers to segment more structures than a 3D representation allows them to do. In fact, the segmentation of more than 25 brain structures into a 3D representation has been achieved by a few works, while using a 2D representation, deep learning works can segment more than 95 brain structures [54].

The use of this type of neural network is not limited to the segmentation of brain tissues or brain structures. These have also been used in the segmentation of brain lesions, as in [55–57], the segmentation of brain tumors [58–60], the detection of ischemic strokes [61], and even genomic prediction of tumors [62]. The most important thing to note from these works is that many use branches within their neural network architectures. In general, they use two branches, where one of them is focused on the extraction of globally related characteristics (global context), while the other is in charge of the extraction of local characteristics (local context) to achieve better segmentation.

One of the architectures most commonly used in medical image segmentation tasks is the U-Net architecture [63]. Due to the structure of its architecture, U-Net has advantages over other convolutional neural network architectures of its time. This was built having a path that encodes the characteristics of the image and then continues with its expansion, that is, an encoder-decoder structure. In addition, to avoid the vanishing gradient and explosion problem, the U-Net architecture incorporates skip connections, between the encoder and decoder layers, which improves performance in small datasets compared with other architectures at the time.

Multiple neural network architectures based on U-Net have been proposed for the field of medical image segmentation. The primary goals of these works were to improve the network by using skip connections between the layers of the coding and expansion path [64,65] and to combine the architecture with others such as SegNet [66]. It is important

to note that several of these studies were also applied to the segmentation of white matter, gray matter, and cerebrospinal fluid from brain magnetic resonance images.

However, convolutional neural networks have serious limitations. One of them is the loss of image characteristics due to pooling operations [67]. This is because the CNNs require these operations to reduce the feature maps resulting from the convolutions and thus reduce the computation required in subsequent layers. Due to this, a large amount of data is necessary in the training process for deep learning networks to be able to generalize and achieve good results [67].

On the other hand, researchers have proposed multiple deep learning architectures based on attention mechanisms such as the Transformer's architecture [22,23,68]. This one was initially proposed in the field of natural language processing [69] as being in charge of transforming one sequence into another from multiple attention blocks [50]. The Transformers replaced the recurrent neural network models (RNN) used until then for the translation of texts because it solved its main weakness. This was because the performance of the recurring models fell when very long sequences were introduced due to the long-term dependency learning problem [68], and although this problem was attacked by the Long Short-Term Memory (LSTM) networks, they did not achieve as good results as the Transformers. This became possible since the latter, through self-attention mechanisms, are capable of processing the entire sequence entered, even if it is very long, optimizing processing times due to parallel processes within the network.

Thus, the scientific community has achieved that the Transformer's architecture can obtain results comparable to those established as the state of the art in computer vision methods [70]. In fact, some methods based on Transformers' architectures were proposed for the segmentation of medical images. The TransUNet [71] network, which is based on the U-Net architecture [63], consists of a set of convolutional layers to extract the most important characteristics of the image. The resulting feature maps are then the input to successive attention blocks, which then send this output to the decoder. The decoder is fabricated of convolutional and upsampling layers to achieve the output of a segmented image. It is necessary to mention that the set of convolutional layers is connected with the layers of the decoder through skip connections.

Also, another Transformer-based architecture is the Medical-Transformer network [72], which is based on the use of two branches for its operation. The important thing to highlight in this study is that it has a local and global branch, as has been proposed in various convolutional neural network architectures and the use of convolutions in the feature coding process. Specifically, the local branch has a greater number of encoder and decoder blocks than the global branch. The encoder blocks are fabricated of 1×1 convolutional layers, normalization, Rectified Linear Unit (ReLU) activations, and multiple layers of attention for its operation, while the decoder has closed axial attention layers.

In this study, a 3D architecture of deep neural networks is proposed for the task of segmenting volumes associated with brain structures from MRI. Our proposal uses an encoder/decoder approach, strengthening the connection between them by incorporating self-attention modules and skip connections. The attention modules as well as the convolution operations allow the network to incorporate global and local characteristics, and achieve a more detailed segmentation of the edges of structures.

2. Material and Methods

2.1. Dataset

We used the publicly available Mindboggle-101 dataset [73]. This dataset contains 101 manually labeled human brain images from healthy patients, containing more than 100 brain structures in each volumetric segmentation file *aseg + aparc*. In this dataset, brain structures are labeled based on the brain hemisphere where they are located (e.g., left hippocampus and right hippocampus).

2.2. Data Preprocessing

We performed some preprocessing steps. We created an array representation of the original MRIs and segmentation masks in the dataset making it straightforward for training deep learning models. Additionally, for the selected structures, we created a label remapping strategy with IDs 1 to 37 having the ID 0 set for background. We created the segmentation ground truth files by taking the aseg+aparc files and mapping the existing IDs to our desired IDs. After this, we applied a min-max scaling to the voxel values to be in a range between 0 and 1. This intensity rescaling was performed using the histogram equalization technique over each individual MRI.

Then, we applied a filter in each MRI where empty slides were removed from the brain volumes, leaving on average 192 slides per brain plane. These preprocessed MRI volumes were divided into nonoverlapping subvolumes of size $64 \times 64 \times 64$ voxels that were saved in single files containing stacks of volumes per brain MRI (see Figure 1). Finally, we divided the dataset into two sets in the ratio of 8:2. The first set was for training the neural network architecture, and the second one was for validating it. Due to the Mindboggle-101 dataset contains multiple datasets such as OASIS-TRT-20 which contains 20 MRIs, NKI-22 that contains 22 MRIs, among others, we performed the dataset division maintaining the original dataset distribution, making sure that the validation set had at least one MRI from all the datasets.

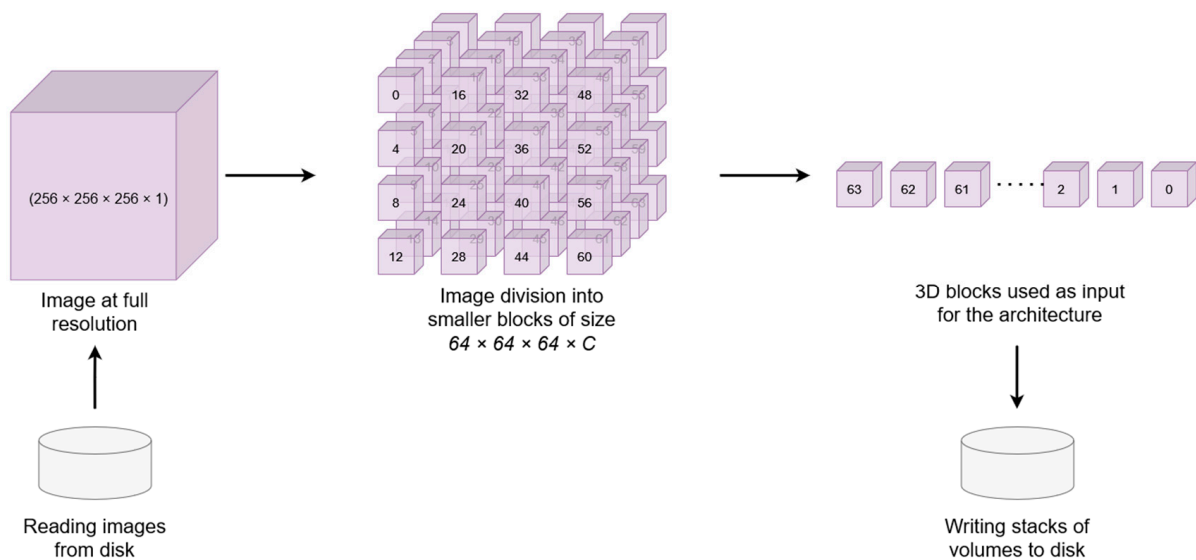


Figure 1. MRI volume divided into nonoverlapping subvolumes of size $64 \times 64 \times 64$ voxels.

2.3. Computational Resources

The computational implementations were performed with the open source library for numerical computation Tensorflow and run on a computer with a 5th generation Intel I7 4820k@3.70 GHz processor, 64 GB of RAM memory, and two Nvidia 1080TI video cards with 11 GB of GDDR 5x RAM at 405 MHz.

2.4. Model for Brain Structures Segmentation from MRI

The proposed deep neural network architecture is structured as an encoder-decoder architecture. The contracting path follows the typical architecture of a convolutional neural network. However, we applied Transformer layers at the end of this path using the extracted feature maps from the CNN layers. The expansive path was composed of a successive combination of convolutional neural networks and upsampling layers in order to reach the original spatial resolution (see Figure 2). To avoid gradient vanishing and explosion problems, we adopted skip connections between the encoder-decoder paths via the usage of Res paths, initially proposed in [66].

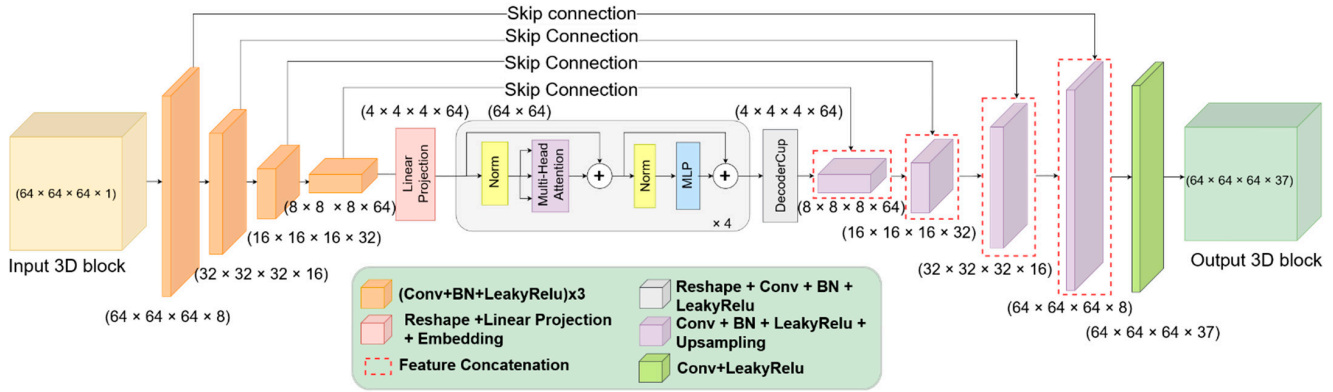


Figure 2. Deep neural network architecture for 3D brain MRI segmentation. The encoder part is composed of convolutional layers and Transformer layers, while the decoder has combinations of convolutional neural networks with upsampling layers.

The architecture input is as follows: given an image $x \in \mathbb{R}^{H \times W \times D \times C}$ where H represents height, W width, D depth, and C the numbers of channels. The objective of this study is to segment multiple anatomical brain structures, predicting the corresponding pixel-wise label maps of size $H \times W \times D$.

2.5. Self-Attention Mechanism for Brain Segmentation

We used self-attention mechanisms via Transformers in the encoder path. This consists of successive I layers of Transformers composed of Multi-Head Self-Attention (MHSA) modules and Multi-Layer Perceptron (MLP) blocks, each preceded by a normalization layer. The MLP blocks use the RELU activation function with a regularization dropout layer.

The attention mechanism was computed in parallel inside each of the heads of the MHSA modules in each transformer using a set of vectors named as query, key, and value vectors [50]. The query vector $q \in \mathbb{R}^d$ was matched against all key vectors organized in a matrix $K \in \mathbb{R}^{k \times d}$ using an algebraic operation known as the dot product. The results were then scaled using a scaling factor $\frac{1}{\sqrt{d_k}}$ and normalized using a *softmax* function to obtain the weights. The attention matrix inside each MHSA head is computed as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where Q , K , and V are matrices representing a set of queries, keys, and values, respectively.

Finally, the results of each head were concatenated and linearly projected into a matrix sequence at the end of the MHSA module. This can be described as:

$$MHSA(Q, K, V) = Concat(head_1, head_2, \dots, head_h)W^O \quad (2)$$

where $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$

where the projections are parameter matrices $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_v}$ and $W^O \in \mathbb{R}^{hd_v \times d_{model}}$ is the dimension of the Transformer's hidden layers.

In the proposed architecture, we used the extracted feature maps from previous convolutional layers as the input of the first transformer layer, using a trainable linear projection. Indeed, we reshaped the feature maps $x \in \mathbb{R}^{H \times W \times D \times C}$ into a flattened representation as the transformer layers expect a sequence as input. Then, we applied positional embedding over the feature maps to add location information for the segmentation process. This can be described as:

$$f x_i^{q,k,v} = FeatureMapsEmbedding(flatten(x_i)) \quad (3)$$

where positional embedding adds location information useful in the segmentation process. This can be seen as:

$$z_0 = [F_1; F_2; F_3; \dots; F_N] + E_{pos}, \quad F \in \mathbb{R}, \quad E_{pos} \in \mathbb{R}^{N \times L} \quad (4)$$

where F denotes the feature maps in conjunction with the linear projection and E_{pos} the position embedding and $N = \frac{H \times W \times D \times C}{16}$. After successive layers of Transformers, the output of the last transformer has a shape $z_l \in \mathbb{R}^{d \times N}$. We applied a reshape before the decoder path to recover its 3D dimensionality.

2.6. Loss Functions and Class Weights

Segmentation of brain structures is a highly imbalanced problem due to the significant differences in size in the structures, presenting greater availability of information in the image for those of greater size. Even the size difference between the structures and the background is usually significant. Therefore, multiple loss functions and weighting strategies for loss functions were proposed for improving imbalanced brain structure segmentation [74]. In the proposed approach, we used a combination of Dice Loss [75] and Focal Loss [76].

Dice Loss (DL) has its origin in the Dice Similarity Coefficient (DSC), which is widely used as a metric for computer vision segmentation to calculate the similarity between two images. Later, in [75], it was adapted as a loss function useful in medical image segmentation tasks, improving the imbalance problem between foreground and background. It is formulated as:

$$\text{Weighted Dice Loss} = 1 - 2 \frac{\sum_{j=1}^S w_j \sum_{i=1}^N y_{ij} p_{ij}}{\sum_{j=1}^S w_j \sum_{i=1}^N y_{ij} + p_{ij}} \quad (5)$$

where w_j is the weight of the j th brain structure and S is the total number of segmentation classes, y_{ij} is the label of voxel i to belong to brain structure j and p_{ij} is the probability of voxel i to belong to brain structure j .

Meanwhile, Focal Loss (FL) is a variation in Binary Cross-Entropy that works better with highly imbalanced datasets. It down-weights the contribution of easy examples and mostly focuses on the hard ones. It can be described as follows:

$$FL(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t) \quad (6)$$

where p_t with $p \in [0, 1]$ is the model's estimated probability for the class, $(1 - p_t)^\gamma$ is the modulating term with γ as the focusing parameter that controls its strength.

In this study, we used a combination of these two functions.

2.7. Metrics

In order to evaluate the performance of the proposed segmentation method, the ground truth and model prediction from the MRIs were compared. The selected metric for this comparison is the DSC which can be seen as a harmonic mean of precision and recall metrics. This metric can be mathematically expressed as:

$$DSC(x, y) = 2 \times \frac{x \cap y}{x + y} \quad (7)$$

where x represents the image ground truth and y represents the predicted output for that image. The marked labels for x and y should be binary represented for each class where all voxels included in a given class should have a value of 1 and 0 for all the others. Therefore, the DSC must be calculated individually for each class having output values in a range between 0 (no segmentation) and 1 (perfect segmentation).

Consequently, *precision* is the accuracy of positive predictions while *recall* is the ratio of positive elements that were predicted correctly by the segmentation model. These metrics can be expressed as:

$$precision = \frac{TP}{TP + FP} \quad (8)$$

$$recall = \frac{TP}{TP + FN} \quad (9)$$

where TP is the number of true positive predictions, FP the number of false positive predictions and FN the number of false negative predictions for a given class.

Also, the Intersection over Union (*IoU*) metric is included in this study as an evaluation metric for specific structures. It is useful for comparing similarities between two shapes A and B and determining true positives and false positives from a set of predictions. It can be expressed as:

$$IoU = \frac{A \cap B}{A \cup B} \quad (10)$$

2.8. Training Process

The training process used a lineal learning rate schedule, initially set at 0.001 and decreased after the 12th iteration to a power of 0.5, while the batch size is set by default at 8. It used the Adam algorithm as the neural network optimizer. For the transformer architecture based on the Visual Transformer (ViT) architecture [77], we set the successive layers and heads per layer at 4, the hidden size at 64, the MLP size at 192, the dropout rate at 0.1, the normalization rate at 0.0001, and a patch resolution of $8 \times 8 \times 8$. It is important to mention that the hyper-parameters were chosen via experimental design.

3. Results

3.1. Performance of Proposed Deep Neural Network Architecture

We quantitatively and visually evaluated the performance of brain structures segmentation.

Figure 3 shows the number of voxels for each of the 37 selected structures from Mindboggle-101 dataset. It can be seen that there are significant differences between classes. To mitigate the effect of class imbalance we use a loss function combining the weighted coefficient Dice and the Focal Loss.

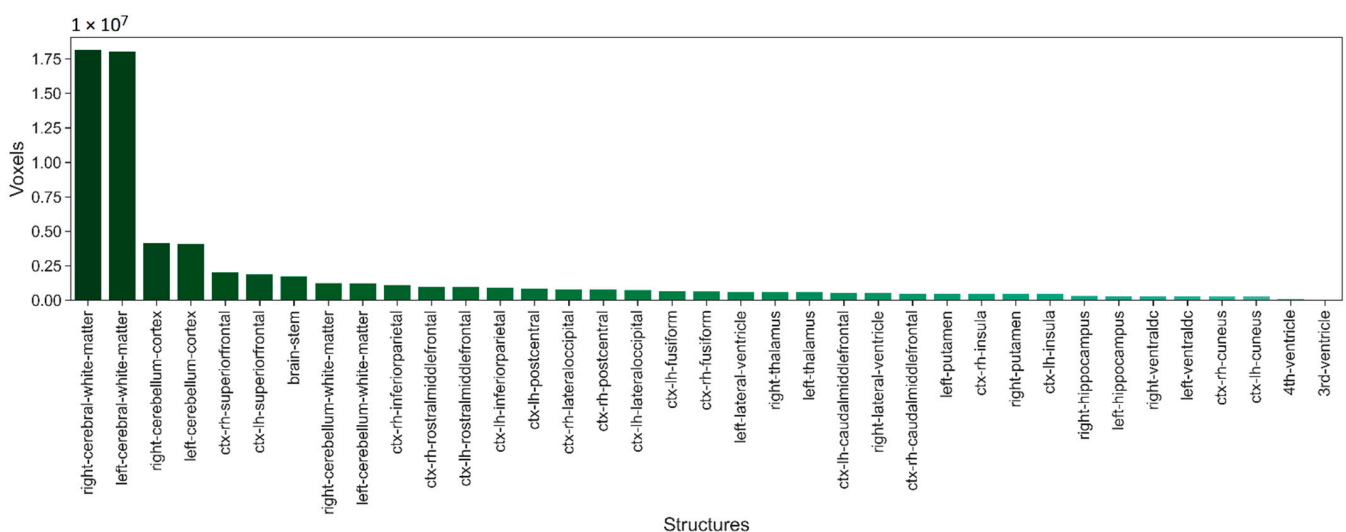


Figure 3. Number of voxels for each of the 37 selected structures from Mindboggle-101 dataset.

The combination of these two loss functions helped us to alleviate the imbalance problem in the segmentation of anatomical brain structures and encourages the correct segmentation of tissue boundaries. Indeed, the use of class weights while training deep

neural network architecture was necessary due to the large number of small structures the brain has compared with the total number of brain voxels. In order to calculate the class weights, we used the median frequency balancing algorithm, which is formulated as follows:

$$\alpha_c = \text{median_freq} / \text{freq}(c) \quad (11)$$

where $\text{freq}(c)$ is the total number of voxels of class c divided by the total number of voxels of the MRIs where class c appeared and median_freq is the median of the calculated frequencies.

A graphic example of the segmentation result of the proposed deep neural network architecture can be seen in Figure 4, where we show the results in the axial, sagittal and coronal planes. The local details at the edges of the structures as well as the global features can be noted compared with the reference image. Quantitatively, we calculated the Precision, Recall, and Dice Score per segmented brain structure (see Table 1). These results show that there are still problems with the segmentation of some structures, mainly small structures that tend to lower values of quality metrics.

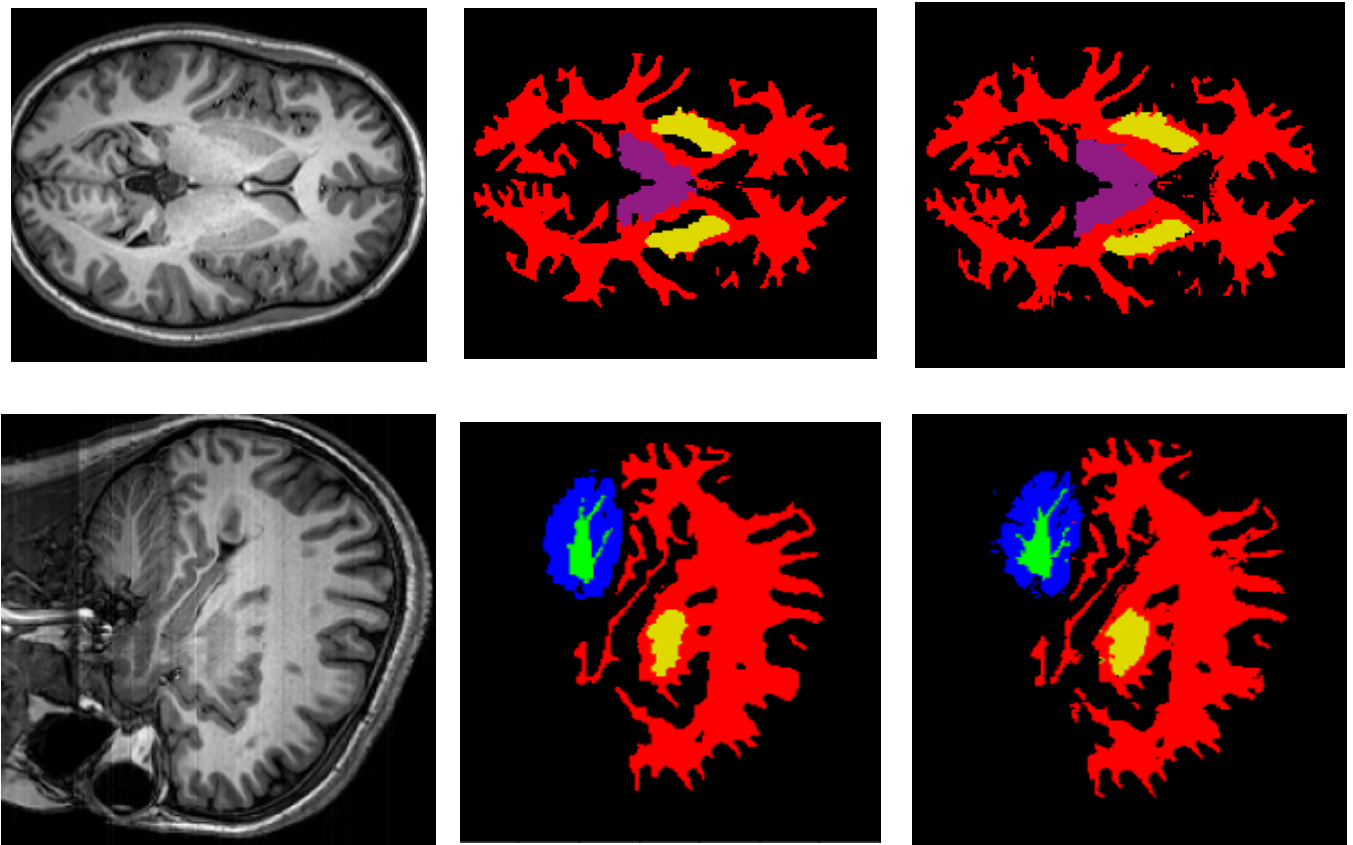


Figure 4. Cont.

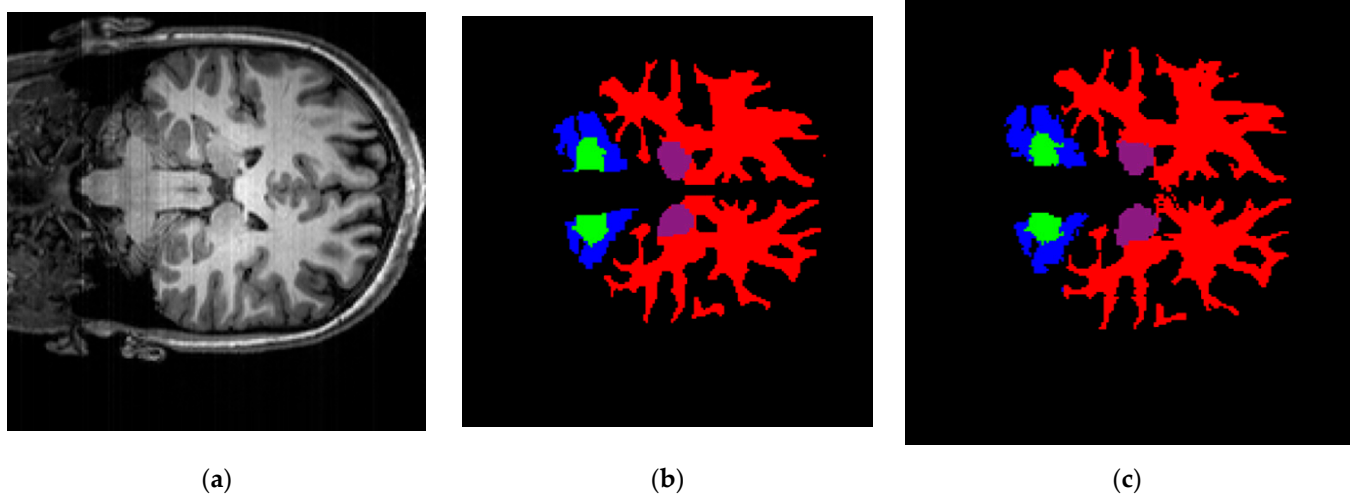


Figure 4. Segmentation results of the proposed architecture in the axial, sagittal and coronal planes where red, green, blue, purple, and yellow colors represent cerebral white matter, cerebellum white matter, cerebellum cortex, thalamus, and putamen structures, respectively. (a) Original MRI slide; (b) ground truth mask of the slide; (c) predicted MRI mask of the slide.

Table 1. Segmentation results per brain structure in a testing MRI.

Brain Structure	Precision	Recall	Dice Score	IoU
Left cerebral white matter	0.95	0.91	0.93	0.86
Right cerebral white matter	0.97	0.89	0.93	0.86
Left cerebellum white matter	0.90	0.75	0.82	0.69
Right cerebellum white matter	0.93	0.77	0.85	0.73
Left cerebellum cortex	0.87	0.82	0.84	0.73
Right cerebellum cortex	0.89	0.72	0.80	0.66
Left lateral ventricle	0.64	0.91	0.75	0.60
Right lateral ventricle	0.78	0.91	0.84	0.72
Left thalamus	0.80	0.92	0.86	0.74
Right thalamus	0.90	0.89	0.89	0.80
Left putamen	0.85	0.84	0.85	0.73
Right putamen	0.91	0.81	0.86	0.75
3rd ventricle	0.57	0.96	0.72	0.56
4th ventricle	0.67	0.94	0.78	0.64
Brain stem	0.87	0.93	0.90	0.83
Left hippocampus	0.88	0.67	0.76	0.62
Right hippocampus	0.89	0.80	0.84	0.73
Left ventral DC	0.78	0.83	0.80	0.68
Right ventral DC	0.62	0.87	0.72	0.57
Ctx left caudal middle frontal	0.84	0.43	0.57	0.40
Ctx right caudal middle frontal	0.50	0.24	0.32	0.20
Ctx left cuneus	0.56	0.65	0.60	0.44
Ctx right cuneus	0.54	0.74	0.62	0.46
Ctx left fusiform	0.68	0.61	0.64	0.48
Ctx right fusiform	0.78	0.65	0.71	0.55
Ctx left inferior parietal	0.64	0.54	0.58	0.42
Ctx right inferior parietal	0.60	0.70	0.65	0.49
Ctx left lateral occipital	0.69	0.74	0.71	0.56
Ctx right lateral occipital	0.73	0.69	0.71	0.56
Ctx left post central	0.54	0.82	0.66	0.49
Ctx right post central	0.71	0.70	0.71	0.55
Ctx right rostral middle frontal	0.57	0.81	0.67	0.50

Table 1. *Cont.*

Brain Structure	Precision	Recall	Dice Score	IoU
Ctx left rostral middle frontal	0.51	0.82	0.63	0.46
Ctx left superior frontal	0.74	0.81	0.77	0.63
Ctx right superior frontal	0.77	0.79	0.78	0.65
Ctx left insula	0.81	0.84	0.82	0.70
Ctx right insula	0.70	0.87	0.78	0.64
Macro average	0.75	0.78	0.75	0.63
Weighted average	0.97	0.97	0.97	0.95

Although to achieve the training process of the deep neural network architecture the brain is divided into blocks of equal size, the results show that the segmented structures maintain spatial coherence and can recover its representative organic form as can be seen in a 3D visual representation shown in Figure 5.

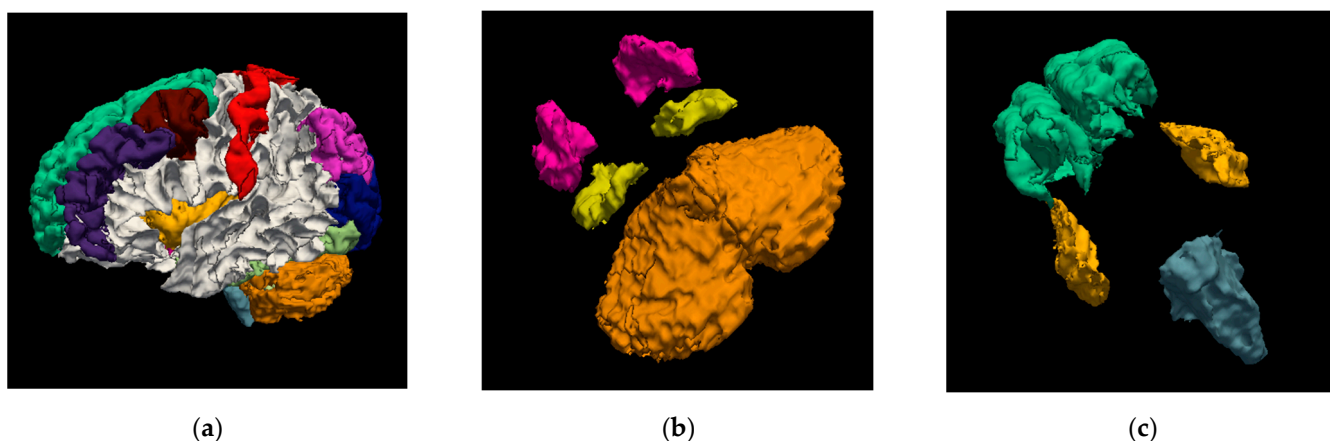


Figure 5. Segmentation results of the proposed architecture in 3D. (a) Segmentation of the 37 brain structures; (b) segmentation of the cerebellum cortex (orange), putamen (magenta), and hippocampus structures (yellow); (c) segmentation of the brain stem (gray), insula (yellow), and superior frontal structures (green).

3.2. Patch Resolution Size Determination

For this experiment, we set four successive layers of transformers with patch resolution blocks of $8 \times 8 \times 8$, hidden size at 64, MLP size at 192; dropout rate at 0.1, and normalization rate at 0.0001. Afterwards, we applied a reshape before the decoder path to recover its 3D dimensionality. Therefore, in order to obtain finer details (local information), it was necessary to use convolutional layers at the beginning of the encoder path following a classical U-Net [63]-based architecture with the proposed architecture.

We experimented with patch resolution sizes related to the transformer layers. As was initially observed in [71], patch resolution size is important since it dictates the number of complex dependencies that each element will have, with others, obtaining finer details in the segmentation process. The ideal case would be to have a patch resolution size of $1 \times 1 \times 1$. Nevertheless, there are not enough computational resources to train a deep neural network architecture based on this patch resolution size. Consequently, we ran experiments on the segmentation of three structures with patch sizes of $16 \times 16 \times 16$ and $8 \times 8 \times 8$ to see its influence on the segmentation of brain structures (see Figure 6).

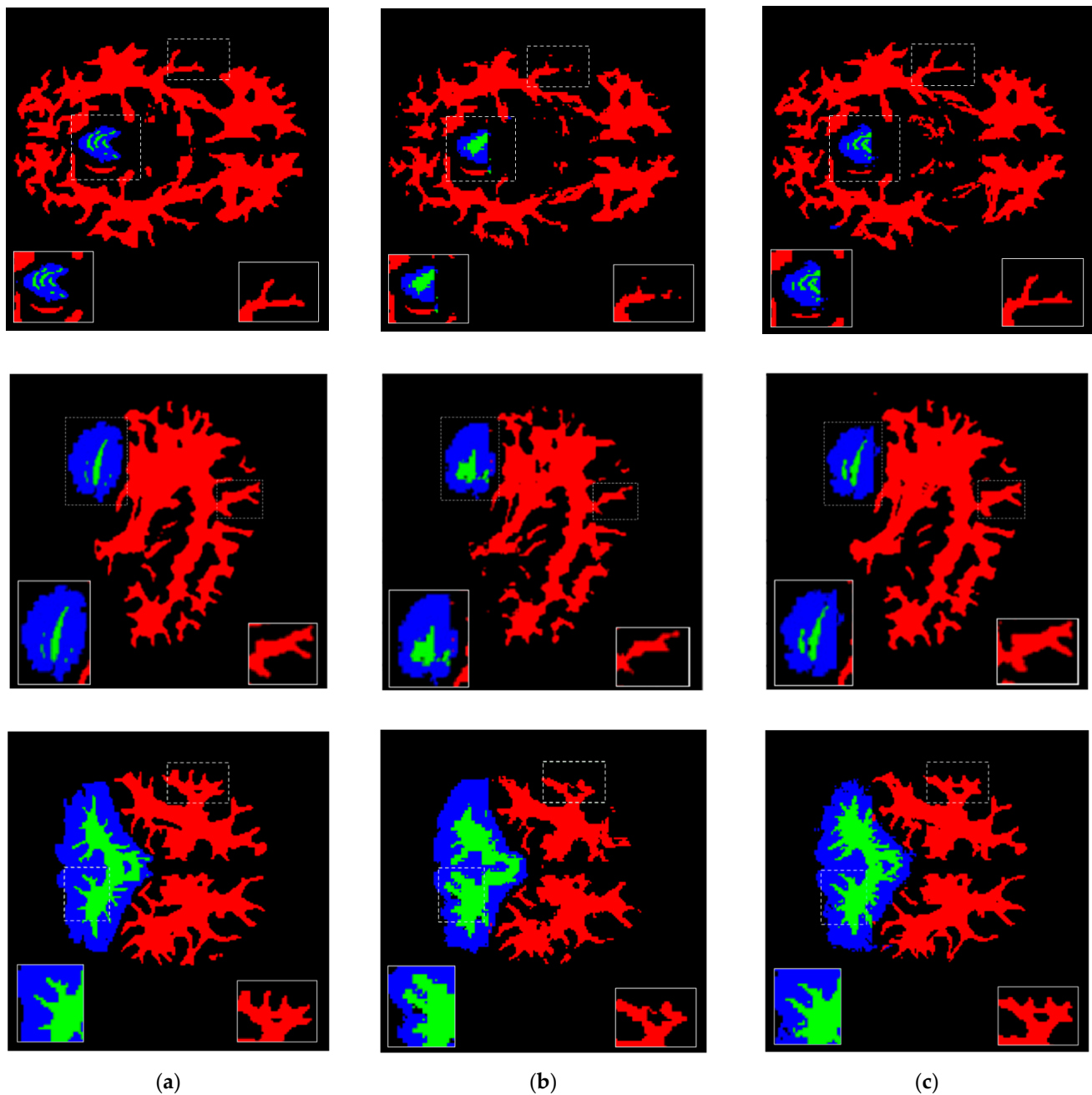


Figure 6. Patch size influence comparison on structure details segmentation in the axial, sagittal and coronal planes where red, green, and blue colors represent cerebral white matter, cerebellum white matter and cerebellum cortex structures, respectively. (a) Ground truth mask; (b) segmentation with patch resolution size of $16 \times 16 \times 16$; (c) segmentation with patch resolution size of $8 \times 8 \times 8$.

3.3. Comparison with Other Methods

At present, the majority of the proposed deep neural network architectures for brain segmentation using Transformers are oriented toward the segmentation of brain tumors. Therefore, it is highly difficult to have a fair comparison since these models were oriented to the segmentation of one class, and not for multiple classes excluding background. Because of this, we implemented the 3D U-Net architecture, using it as our baseline, with an identical experimental setup. This comparison using the Dice score and the Wilcoxon signed-rank test can be seen in Table 2.

Table 2. Comparison between methods by the Dice Score and p -value for the Wilcoxon signed-rank test comparing proposed-UNet, proposed-DenseUNet samples pairs using the Mindboggle-101 dataset.

Model	Brain Structures	Mean Dice Score	p -Value
UNet (baseline)	37	0.790 ± 0.0210	0.0012850
DenseUNet (finetuned)	102	0.819 ± 0.0110	0.0211314
Proposed model	37	0.900 ± 0.0360	-

The time needed to perform the segmentation by this architecture and the comparison with other deep learning models is shown in Table 3. It is important to mention that transformer layers, thanks to the self-attention mechanism, are capable of processing entire sequences in parallel, optimizing processing times. Unlike CPU processing units, the GPU architecture was specifically designed to process data in parallel, allowing the proposed model to take full advantage of computational resources and the Transformer's processing pipeline.

Table 3. Segmentation time per brain structure for a single MRI scan.

Model	Segments	Time Per Brain Structure	Mean Dice Score
DeepNAT [78]	27	~133 s (on a Multi-GPU Machine)	0.906
QuickNAT [79]	27	~0.74 s (on a Multi-GPU Machine)	0.901
DenseUNet	102	0.64 s (± 0.0091 s) (Single GPU Machine)	0.819
FreeSurfer [79]	~190	~75.8 s	-
Proposed model	37	0.038 s (± 0.0016 s) (on a Multi-GPU Machine)	0.903

4. Discussion and Future Work

This study presents a deep learning-based model for the segmentation of 37 brain structures using transformer models. This network was trained with the manually annotated dataset Mindboggle-101, which contains 101 MRIs with its respective segmentation files processed using the Desikan-Killiany-Tourville (DKT) protocol [73]. In the scientific community, it is common to find multiple approaches to perform the segmentation of MRIs. Therefore, this architecture was indirectly trained to perform segmentation based on the DKT protocol due to the used dataset.

Our architecture includes self-attention modules to strengthen the connection between the encoding and decoding phases based on convolutional neural networks. The capabilities of self-attention modules add to the model the possibility of retaining features across voxels in the input patches of the model. Unlike 2D-based models, the 3D architecture can find voxel relationships in the three different planes, maximizing the use of the spatial nature implicit in MRI.

In addition, the results of the proposed segmentation model show that the quality metrics have a wide range of values. For the Dice Score, for example, the values vary from 0.32 to values of 0.93, showing low-quality segmentations for some structures. We find that the lower values tend to be related to structures with smaller volumes. The geometry at the edges of the structures is a factor that we consider influences the quality of the segmentation. Structures with borders of highly variable geometry tend to have segmentations with more error. Simpler edge structures generally result in more stable quality segmentations. Our intuition in this regard is that this is due, mainly, to the class imbalance problem and lack of enough data to train the model for those structures specifically. For this reason, it is important in the future to explore other methods that allow addressing this problem, for example, improving the calculation of the weights of the classes used in the loss functions similar to what is performed in [74], or using additional data augmentation techniques to increase the samples of classes with less information. Another factor that we considered in the analysis is the fact that deep learning methods based on transformers lack the inductive biases inherent in CNNs requiring large amounts of data to be able to generalize

well [77], so their usage in small-size medical datasets remains difficult without any internal modification in their self-attention module. Incorporating these modifications can allow us to improve the segmentation of a large number of highly unbalanced brain structures using a 3D approach.

The patch resolution size is a determining factor to obtain finer details at the edges of segmentations. The experiments show an inverse relationship with size, that is, the smaller the patch size, the more detailed segmentation is obtained at the edges; the larger the patch size, the segmentations tend to be less detailed. It must be considered not all structures have geometrically complex edges, there are structures with simpler geometrics. Therefore, a trade-off between the computational cost of reducing the patch size and the more detailed segmentation requirements must be considered. Given the 3D representation used in this study and the memory requirements, it was not possible to explore values smaller than $8 \times 8 \times 8$.

The results show that our method uses less time for segmentation with a Mean Dice Score similar to those found in the state of the art. Additionally, the segmentation of more than 25 brain structures into a 3D representation is a difficult task that has only been reported by a few groups of authors [54] due to computational and memory limitations. However, it is not competitive in terms of the number of segments where the latest 2D deep learning-based approaches are able to segment more than 100 structures. Consequently, further study should be carried out to optimize the use of memory and computational resources in the proposed architecture to segment more brain structures with a strong focus on the transformer architecture.

Our method still has deficiencies related to the variation in the segmentation quality for different structures. Class imbalance, as well as the broad geometric nature of the edges are factors for which our method is still sensitive. The number of segmented structures is also a limitation, it is desirable to be able to segment a greater number of structures, especially compared with 2D-based approaches.

In future work, we will explore the existing computational and memory limitations in our proposed architecture with a high focus on the transformer layers to see whether a different tokenization of the patched feature maps can improve its performance and segment more brain structures.

Author Contributions: Conceptualization, C.L.-B. and G.S.-T.; methodology, C.L.-B., G.S.-T. and J.B.-B.; software, C.L.-B.; validation, G.S.-T., J.B.-B. and C.L.-B.; formal analysis, C.L.-B.; investigation, C.L.-B.; resources, G.S.-T. and J.B.-B.; data curation, C.L.-B. and G.S.-T.; writing—original draft preparation, C.L.-B. and G.S.-T.; writing—review and editing, G.S.-T. and J.B.-B.; visualization, G.S.-T.; supervision, G.S.-T. and J.B.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universidad Nacional de Colombia, Sede Medellín, grant number BEDA2021II.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Miller-Thomas, M.M.; Benzinger, T.L.S. Neurologic Applications of PET/MR Imaging. *Magn. Reson. Imaging Clin. N. Am.* **2017**, *25*, 297–313. [[CrossRef](#)] [[PubMed](#)]
2. Mier, W.; Mier, D. Advantages in functional imaging of the brain. *Front. Hum. Neurosci.* **2015**, *9*, 249. [[CrossRef](#)] [[PubMed](#)]
3. Neeb, H.; Zilles, K.; Shah, N.J. Fully-automated detection of cerebral water content changes: Study of age- and gender-related H₂O patterns with quantitative MRI. *NeuroImage* **2006**, *29*, 910–922. [[CrossRef](#)] [[PubMed](#)]
4. Shenton, M.E.; Dickey, C.C.; Frumin, M.; McCarley, R.W. A review of MRI findings in schizophrenia. *Schizophr. Res.* **2001**, *49*, 1–52. [[CrossRef](#)]
5. Widmann, G.; Henninger, B.; Kremser, C.; Jaschke, W. MRI Sequences in Head & Neck Radiology—State of the Art. *ROFO Fortschr. Geb. Rontgenstr. Nuklearmed.* **2017**, *189*, 413–422. [[CrossRef](#)]

6. Yu, H.; Yang, L.T.; Zhang, Q.; Armstrong, D.; Deen, M.J. Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives. *Neurocomputing* **2021**, *444*, 92–110. [[CrossRef](#)]
7. Xie, X.; Niu, J.; Liu, X.; Chen, Z.; Tang, S.; Yu, S. A survey on incorporating domain knowledge into deep learning for medical image analysis. *Med. Image Anal.* **2021**, *69*, 101985. [[CrossRef](#)]
8. Ilhan, U.; Ilhan, A. Brain tumor segmentation based on a new threshold approach. *Procedia Comput. Sci.* **2017**, *120*, 580–587. [[CrossRef](#)]
9. Deng, W.; Xiao, W.; Deng, H.; Liu, J. MRI brain tumor segmentation with region growing method based on the gradients and variances along and inside of the boundary curve. In Proceedings of the 2010 3rd International Conference on Biomedical Engineering and Informatics, Yantai, China, 16–18 October 2010; Volume 1, pp. 393–396.
10. Ashburner, J.; Friston, K.J. Unified segmentation. *NeuroImage* **2005**, *26*, 839–851. [[CrossRef](#)]
11. Liu, J.; Guo, L. A New Brain MRI Image Segmentation Strategy Based on K-means Clustering and SVM. In Proceedings of the 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics, Hangzhou, China, 26–27 August 2015; Volume 2, pp. 270–273.
12. Zhao, X.; Zhao, X.-M. Deep learning of brain magnetic resonance images: A brief review. *Methods* **2021**, *192*, 131–140. [[CrossRef](#)]
13. Despotović, I.; Goossens, B.; Philips, W. MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. *Comput. Math. Methods Med.* **2015**, *2015*, e450341. [[CrossRef](#)] [[PubMed](#)]
14. Fischl, B.; Salat, D.H.; Busa, E.; Albert, M.; Dieterich, M.; Haselgrove, C.; van der Kouwe, A.; Killiany, R.; Kennedy, D.; Klaveness, S.; et al. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **2002**, *33*, 341–355. [[CrossRef](#)]
15. Shattuck, D.W.; Leahy, R.M. BrainSuite: An automated cortical surface identification tool. *Med. Image Anal.* **2002**, *6*, 129–142. [[CrossRef](#)]
16. Jenkinson, M.; Beckmann, C.F.; Behrens, T.E.J.; Woolrich, M.W.; Smith, S.M. FSL. *NeuroImage* **2012**, *62*, 782–790. [[CrossRef](#)]
17. Yushkevich, P.A.; Gao, Y.; Gerig, G. ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 3342–3345.
18. Pieper, S.; Halle, M.; Kikinis, R. 3D Slicer. In Proceedings of the 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821), Arlington, VA, USA, 15–18 April 2004; Volume 1, pp. 632–635.
19. Please Help Sustain the Horos Project—Horos Project. Available online: <https://horosproject.org/download-donation/> (accessed on 24 August 2021).
20. Qin, C.; Wu, Y.; Liao, W.; Zeng, J.; Liang, S.; Zhang, X. Improved U-Net3+ with stage residual for brain tumor segmentation. *BMC Med. Imaging* **2022**, *22*, 14. [[CrossRef](#)]
21. Sun, J.; Peng, Y.; Guo, Y.; Li, D. Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3D FCN. *Neurocomputing* **2021**, *423*, 34–45. [[CrossRef](#)]
22. Dai, W.; Woo, B.; Liu, S.; Marques, M.; Tang, F.; Crozier, S.; Engstrom, C.; Chandra, S. Can3d: Fast 3d Knee Mri Segmentation Via Compact Context Aggregation. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 13–16 April 2021; pp. 1505–1508.
23. Deepthi Murthy, T.S.; Sadashivappa, G. Brain tumor segmentation using thresholding, morphological operations and extraction of features of tumor. In Proceedings of the 2014 International Conference on Advances in Electronics Computers and Communications, Bangalore, India, 10–11 October 2014; pp. 1–6.
24. Polakowski, W.E.; Cournoyer, D.A.; Rogers, S.K.; DeSimio, M.P.; Ruck, D.W.; Hoffmeister, J.W.; Raines, R.A. Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussians and derivative-based feature saliency. *IEEE Trans. Med. Imaging* **1997**, *16*, 811–819. [[CrossRef](#)]
25. Wani, M.A.; Batchelor, B.G. Edge-region-based segmentation of range images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 314–319. [[CrossRef](#)]
26. Wu, J.; Ye, F.; Ma, J.-L.; Sun, X.-P.; Xu, J.; Cui, Z.-M. The Segmentation and Visualization of Human Organs Based on Adaptive Region Growing Method. In Proceedings of the 2008 IEEE 8th International Conference on Computer and Information Technology Workshops, Washington, WA, USA, 8–11 July 2008; pp. 439–443.
27. Passat, N.; Ronse, C.; Baruthio, J.; Armspach, J.-P.; Maillot, C.; Jahn, C. Region-growing segmentation of brain vessels: An atlas-based automatic approach. *J. Magn. Reson. Imaging JMRI* **2005**, *21*, 715–725. [[CrossRef](#)]
28. Gibbs, P.; Buckley, D.L.; Blackband, S.J.; Horsman, A. Tumour volume determination from MR images by morphological segmentation. *Phys. Med. Biol.* **1996**, *41*, 2437–2446. [[CrossRef](#)]
29. Pohlman, S.; Powell, K.A.; Obuchowski, N.A.; Chilcote, W.A.; Grundfest-Broniatowski, S. Quantitative classification of breast tumors in digitized mammograms. *Med. Phys.* **1996**, *23*, 1337–1345. [[CrossRef](#)] [[PubMed](#)]
30. Maksoud, E.A.A.; Elmogy, M.; Al-Awadi, R.M. MRI Brain Tumor Segmentation System Based on Hybrid Clustering Techniques. In Proceedings of the Advanced Machine Learning Technologies and Applications, Cairo, Egypt, 28–30 November 2014; Hassanien, A.E., Tolba, M.F., Taher Azar, A., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2014; pp. 401–412.
31. Artaechevarria, X.; Munoz-Barrutia, A.; Ortiz-de-Solorzano, C. Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data. *IEEE Trans. Med. Imaging* **2009**, *28*, 1266–1277. [[CrossRef](#)] [[PubMed](#)]

32. Coupé, P.; Manjón, J.V.; Fonov, V.; Pruessner, J.; Robles, M.; Collins, D.L. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* **2011**, *54*, 940–954. [[CrossRef](#)] [[PubMed](#)]
33. Wang, H.; Suh, J.W.; Das, S.R.; Pluta, J.B.; Craige, C.; Yushkevich, P.A. Multi-Atlas Segmentation with Joint Label Fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 611–623. [[CrossRef](#)]
34. Wu, G.; Wang, Q.; Zhang, D.; Nie, F.; Huang, H.; Shen, D. A generative probability model of joint label fusion for multi-atlas based brain segmentation. *Med. Image Anal.* **2014**, *18*, 881–890. [[CrossRef](#)]
35. Kass, M.; Witkin, A.; Terzopoulos, D. Snakes: Active contour models. *Int. J. Comput. Vis.* **1988**, *1*, 321–331. [[CrossRef](#)]
36. Wu, Z.; Guo, Y.; Park, S.H.; Gao, Y.; Dong, P.; Lee, S.-W.; Shen, D. Robust brain ROI segmentation by deformation regression and deformable shape model. *Med. Image Anal.* **2018**, *43*, 198–213. [[CrossRef](#)]
37. Rajendran, A.; Dhanasekaran, R. Fuzzy Clustering and Deformable Model for Tumor Segmentation on MRI Brain Image: A Combined Approach. *Procedia Eng.* **2012**, *30*, 327–333. [[CrossRef](#)]
38. Khotanlou, H.; Atif, J.; Colliot, O.; Bloch, I. 3D Brain Tumor Segmentation Using Fuzzy Classification and Deformable Models. In Proceedings of the Fuzzy Logic and Applications, Crema, Italy, 15–17 September 2005; Bloch, I., Petrosino, A., Tettamanzi, A.G.B., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 312–318.
39. Le, W.T.; Maleki, F.; Romero, F.P.; Forghani, R.; Kadoury, S. Overview of Machine Learning: Part 2: Deep Learning for Medical Image Analysis. *Neuroimaging Clin. N. Am.* **2020**, *30*, 417–431. [[CrossRef](#)]
40. Liu, X.; Wang, H.; Li, Z.; Qin, L. Deep learning in ECG diagnosis: A review. *Knowl. Based Syst.* **2021**, *227*, 107187. [[CrossRef](#)]
41. Nogales, A.; García-Tejedor, Á.J.; Monge, D.; Vara, J.S.; Antón, C. A survey of deep learning models in medical therapeutic areas. *Artif. Intell. Med.* **2021**, *112*, 102020. [[CrossRef](#)] [[PubMed](#)]
42. Tng, S.S.; Le, N.Q.K.; Yeh, H.-Y.; Chua, M.C.H. Improved Prediction Model of Protein Lysine Crotonylation Sites Using Bidirectional Recurrent Neural Networks. *J. Proteome Res.* **2022**, *21*, 265–273. [[CrossRef](#)] [[PubMed](#)]
43. Le, N.Q.K.; Ho, Q.-T. Deep transformers and convolutional neural network in identifying DNA N6-methyladenine sites in cross-species genomes. *Methods* **2021**. [[CrossRef](#)] [[PubMed](#)]
44. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [[CrossRef](#)]
45. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [[CrossRef](#)]
46. Bank, D.; Koenigstein, N.; Giryas, R. Autoencoders. *arXiv* **2020**, arXiv:2003.05991.
47. Montúfar, G. Restricted Boltzmann Machines: Introduction and Review. *arXiv* **2018**, arXiv:1806.07066.
48. Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *arXiv* **2018**, arXiv:1808.03314. [[CrossRef](#)]
49. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [[CrossRef](#)]
50. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
51. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **1980**, *36*, 193–202. [[CrossRef](#)] [[PubMed](#)]
52. Nie, D.; Wang, L.; Gao, Y.; Shen, D. Fully convolutional networks for multi-modality iso-intense infant brain image segmentation. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016. [[CrossRef](#)]
53. Bao, S.; Chung, A.C.S. Multi-scale structured CNN with label consistency for brain MR image segmentation. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2018**, *6*, 113–117. [[CrossRef](#)]
54. Henschel, L.; Conjeti, S.; Estrada, S.; Diers, K.; Fischl, B.; Reuter, M. FastSurfer—A fast and accurate deep learning based neuroimaging pipeline. *NeuroImage* **2020**, *219*, 117012. [[CrossRef](#)] [[PubMed](#)]
55. Brosch, T.; Tang, L.Y.W.; Yoo, Y.; Li, D.K.B.; Traboulsee, A.; Tam, R. Deep 3D Convolutional Encoder Networks With Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation. *IEEE Trans. Med. Imaging* **2016**, *35*, 1229–1239. [[CrossRef](#)]
56. Valverde, S.; Cabezas, M.; Roura, E.; González-Villà, S.; Pareto, D.; Vilanova, J.; Ramió-Torrentà, L.; Rovira, À.; Oliver, A.; Lladó, X. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *NeuroImage* **2017**, *155*, 159–168. [[CrossRef](#)]
57. Gabr, R.E.; Coronado, I.; Robinson, M.; Sujit, S.J.; Datta, S.; Sun, X.; Allen, W.J.; Lublin, F.D.; Wolinsky, J.S.; Narayana, P.A. Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: A large-scale study. *Mult. Scler. Houndmills Basingstoke Engl.* **2020**, *26*, 1217–1226. [[CrossRef](#)]
58. Havaei, M.; Davy, A.; Warde-Farley, D.; Biard, A.; Courville, A.; Bengio, Y.; Pal, C.; Jodoin, P.-M.; Larochelle, H. Brain tumor segmentation with Deep Neural Networks. *Med. Image Anal.* **2017**, *35*, 18–31. [[CrossRef](#)]

59. Havaei, M.; Dutil, F.; Pal, C.; Larochelle, H.; Jodoin, P.-M. A Convolutional Neural Network Approach to Brain Tumor Segmentation. In Proceedings of the Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Munich, Germany, 5 October 2015; Crimi, A., Menze, B., Maier, O., Reyes, M., Handels, H., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 195–208.
60. McKinley, R.; Jungo, A.; Wiest, R.; Reyes, M. Pooling-Free Fully Convolutional Networks with Dense Skip Connections for Semantic Segmentation, with Application to Brain Tumor Segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*; Crimi, A., Bakas, S., Kuijff, H., Menze, B., Reyes, M., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 169–177.
61. Chen, L.; Bentley, P.; Rueckert, D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *NeuroImage Clin.* **2017**, *15*, 633–643. [[CrossRef](#)]
62. Akkus, Z.; Ali, I.; Sedlar, J.; Kline, T.L.; Agrawal, J.P.; Parney, I.F.; Giannini, C.; Erickson, B.J. Predicting 1p19q Chromosomal Deletion of Low-Grade Gliomas from MR Images using Deep Learning. *arXiv* **2016**, arXiv:1611.06939.
63. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
64. Kumar, P.; Nagar, P.; Arora, C.; Gupta, A. U-Segnet: Fully Convolutional Neural Network Based Automated Brain Tissue Segmentation Tool. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 3503–3507.
65. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 1856–1867. [[CrossRef](#)]
66. Ibtehaz, N.; Rahman, M.S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Netw.* **2020**, *121*, 74–87. [[CrossRef](#)] [[PubMed](#)]
67. Kwabena Patrick, M.; Felix Adekoya, A.; Abra Mighty, A.; Edward, B.Y. Capsule Networks—A survey. *J. King Saud Univ. Comput. Inf. Sci.* **2019**, *34*, 1295–1310. [[CrossRef](#)]
68. Salehinejad, H.; Baarbe, J.; Sankar, S.; Barfett, J.; Colak, E.; Valaee, S. Recent Advances in Recurrent Neural Networks. *arXiv* **2018**, arXiv:1801.01078.
69. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:1810.04805.
70. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.S.; et al. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. *arXiv* **2021**, arXiv:2012.15840.
71. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
72. Valanarasu, J.M.J.; Oza, P.; Hacihaliloglu, I.; Patel, V.M. Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.10662.
73. Klein, A.; Tourville, J. 101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol. *Front. Neurosci.* **2012**, *6*, 171. [[CrossRef](#)]
74. Sugino, T.; Kawase, T.; Onogi, S.; Kin, T.; Saito, N.; Nakajima, Y. Loss Weightings for Improving Imbalanced Brain Structure Segmentation Using Fully Convolutional Networks. *Healthcare* **2021**, *9*, 938. [[CrossRef](#)]
75. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Jorge Cardoso, M. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Cardoso, M.J., Arbel, T., Carneiro, G., Syeda-Mahmood, T., Tavares, J.M.R.S., Moradi, M., Bradley, A., Greenspan, H., Papa, J.P., Madabhushi, A., et al., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 240–248.
76. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007.
77. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929.
78. Wachinger, C.; Reuter, M.; Klein, T. DeepNAT: Deep Convolutional Neural Network for Segmenting Neuroanatomy. *NeuroImage* **2018**, *170*, 434–445. [[CrossRef](#)] [[PubMed](#)]
79. Roy, A.G.; Conjeti, S.; Navab, N.; Wachinger, C. QuickNAT: A Fully Convolutional Network for Quick and Accurate Segmentation of Neuroanatomy. *arXiv* **2018**, arXiv:1801.04161.