Improved Alzheimer's disease (AD) Diagnosis with U-Net Variants Trained on Brain MRI Scans

Jiaxin Dong A20499780
Toluwani Oyewusi A20425821

CS512 Computer Vision

Abstract

In this project, we use U-Net variations trained on brain magnetic resonance imaging (MRI) scans to offer an improved framework for diagnosing Alzheimer's disease (AD). We investigate a number of cutting-edge U-Net topologies, including U-Net++, with the goal of improving the precision and robustness of AD diagnosis through sophisticated segmentation methods. To further improve the performance of our models in segmenting brain MRI data, we employ a variety of loss functions, including Binary Cross-Entropy (BCE) and Dice Loss. Our experimental findings show that those U-Net topologies have the ability to precisely identify and delineate important AD-related features in brain MRI scans, improving diagnostic performance.

1 Problem Statement

The most common form of dementia, Alzheimer's disease (AD), affects millions of people worldwide and affects patients, caregivers, and healthcare systems. The frequency of AD is anticipated to rise as the world's population ages, highlighting the need for precise and effective diagnostic techniques to identify and treat the condition in its earliest stages [1]. A key region in the early stages and development of AD is the hippocampus, a part of the brain involved in memory formation and retrieval [2][3]. The hippocampus is one of the first brain regions to be impacted by AD, according to studies, making it a strong target to detect early indications of the illness [4]. Magnetic resonance imaging (MRI) has the ability to evaluate and visualize this particular structure of the brain, which can provide intimate details that can lead to the stratification of AD severity. Manually interpreting these photos, however, can take a lot of effort and be prone to mistakes. Recent developments in deep learning and computer vision opened up new possibilities for the creation of automated systems that can evaluate medical images more accurately and effectively. These methods could has the potential to fundamentally alter the way AD is diagnosed, enabling earlier intervention and creating better patient outcomes. Despite the great potential of deep learning in medical imaging, more effort is still required to create and improve models specifically suited to the complexity of diagnosing AD.

The purpose of this project is to provide a deep learning-based method for the automatic detection and diagnosis of AD. U-Net[5], a cutting-edge convolutional neural network (CNN) architecture, and its variations, such as U-Net++ [6], can utilize preprocessed dataset to achieve just that. To ensure the dependability and robustness of the models, the suggested method would comprise of preprocessing and preparing the dataset for training and evaluation with a focus on standardization, normalization, and data augmentation techniques. These models will be created to separate the hippocampus and other regions of interest in

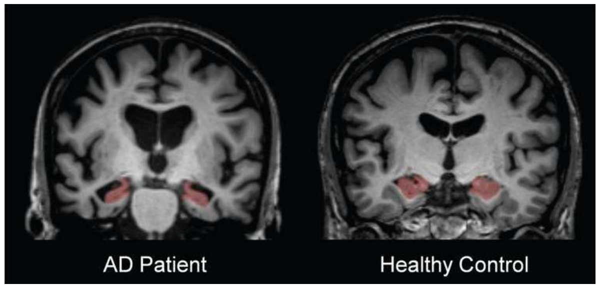

Figure 1: An example image of the hippocampus control. [7]

the brain MRI scans, and then classify the existence of AD using the segmented features. The U-Net architecture and its variations are good candidates for this application since they have demonstrated promising results in a number of medical picture segmentation tasks. The findings of this work will aid in the continued development of automated AD diagnosis tools and shed light on the possible advantages of analyzing brain MRI scans for AD using deep learning methods.

2 Relative Work

MRI scans of the hippocampus provides valuable information in diagnosing AD. Accurate segmentation of the hippocampus, which is essential for storing and retrieving memories and one of the first regions of the brain to be impacted by AD, can shed light on the early recognition and development of the illness [8]. The model can examine the structural alterations and patterns of atrophy unique to AD by focusing on the hippocampus, perhaps providing a more precise and reliable diagnosis. Furthermore, the segmentation of the hippocampus enables clinicians to better track the course of the disease and base their decisions on the structural changes in this important brain region, which can help in the creating of individualized treatment plans.

The U-Net architecture has significantly increased in popularity in the field of medical picture segmentation as a result of its outstanding performance in numerous applications. The majority of the literature on MRI-based medical image segmentation makes use of various iterations of the U-Net architecture, according to the 2022 conference of the Medical Image Computing and Computer Assisted Intervention (MICCAI) Society. U-Net's capacity to capture both high-level and low-level features in images, which is especially helpful when working with complex structures like the brain, can be credited for its effectiveness in medical image segmentation. The encoder of the U-Net architecture captures the contextual information, while the decoder reconstructs the segmented image. This architecture is modeled after a CNN. Additionally, the architecture has skip links between the encoder and decoder layers that enable the model to keep track of fine-grained data and raise the output segmentation quality.

3 Proposed Solution

In our research, we propose an improved method to diagnose AD using U-Net (Fig 2) and U-Net++ (Fig 3) architectures trained on brain MRI scans. The U-Net model is a widely-used architecture for medical image segmentation tasks. It consists of an encoder-decoder structure where the encoder gradually down-samples the input image gradually into abstract feature representations and the decoder up-samples these representations to recover the original image size. The encoder and decoder layers of the model use skip connections, which aid in the preservation of spatial information and make it easier to reconstruct fine-grained details in the segmented output. To make segmentation networks more resistant to minor disturbances like picture translation and rotation, downsampling and upsampling are crucial. Down-sampling increases the receptive field size, minimizes the likelihood of overfitting, and reduces computational complexity, allowing the model to capture more abstract and complicated data. On the other hand, upsampling decodes the abstract characteristics back to the original image size to provide the segmentation outcome.

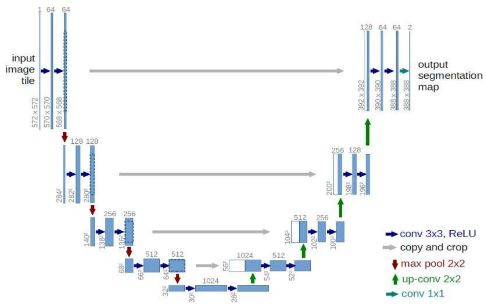

Figure 2: U-Net Model

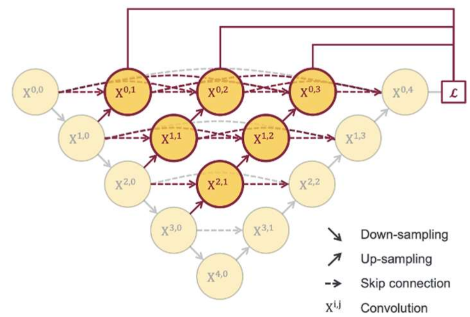

Figure 3: U-Net++ Model

However, the original U-Net architecture has limitations in accommodating the versatility required for various segmentation tasks. We can demonstrate the flexibility and efficiency of the U-Net and U-Net++ models in processing various forms of input data by using cell segmentation as an example. In this instance, we'll make use of a dataset of 354 examples of cell pictures from VisionGate [9], each measuring 96x96 pixels. For the U-Net model, we can evaluate its performance on this cell segmentation task by training it with different numbers of layers (Fig 4). In this example, we can report the following results for the U-Net model with varying depths, Intersection over union (IoU) is used as the metric for comparison. U-Net L1 (1 layer): 88.58%, U-Net L2 (2 layers): 89.39%, U-Net L3 (3 layers): 90.14%, U-Net L4 (4 layers): 88.73% (Fig 5). Applying a 4-layer U-Net model is not essential because cell segmentation is a reasonably straightforward process and there is little performance gain over shallower models. Using a 3-layer U-Net model in this situation seems to provide the best results with the least amount of complexity.

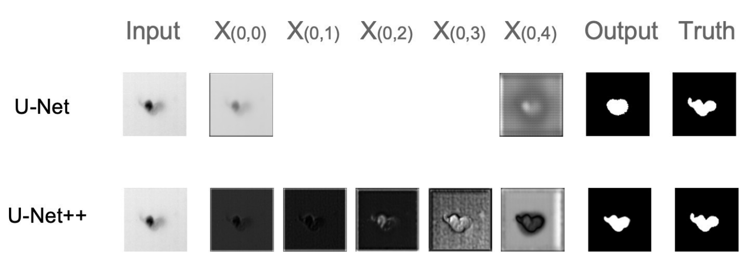

Figure 4: Cell segmentation with 4 layers structure

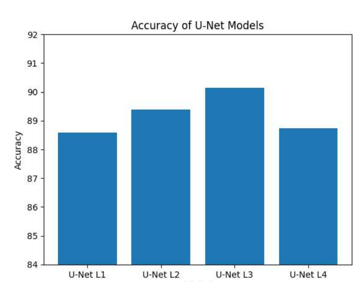

Figure 5: Accuracy

U-Net++ model can be used to address this, which introduces nested U-Net subnetworks at various depths, improving the network's adaptability and efficiency for a wider range of tasks.

To improve upon the capabilities of the original U-Net, U-Net++ uses a layered, densely connected, and extensively supervised structure. Redesigning skip paths to increase feature fusion and densely connecting feature maps for improved information flow are the main goals of the suggested improvements. To improve the efficiency of succeeding layers' use of intermediate representations, concatenation is utilized in place of summation. Deep supervision (Fig 6) is incorporated into the model to improve learning and enable the optimization of loss functions at various scales. This improves training effectiveness by facilitating gradient flow and lowering the chance of vanishing gradients. Pruning is another feature of U-Net++ that offers a speed-accuracy trade-off. The model can achieve quicker inference times while retaining tolerable accuracy levels by pruning less important subnetworks.

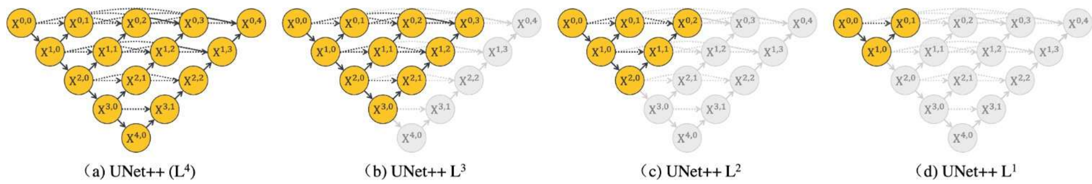

Figure 6: U-Net++ with deep supervision

4 Implementation

I. Prepare Dataset

In this section, we review how to use brain MRI images to diagnose AD using U-Net and U-Net++ models. The dataset used in this study is composed of MRI segmentations of the hippocampus, and obtained from Kaggle's MRI Hippocampus Segmentation competition (Fig 7). It includes images from 135 patients, with 100 designated for training and 35 for testing. All cases in the dataset exhibit Alzheimer's disease. The dataset contains 18,711 images instead of the expected 18,900 (189*100) due to a missing label folder for one patient's hippocampus data. Out of these images, 2,961 contain the hippocampus.

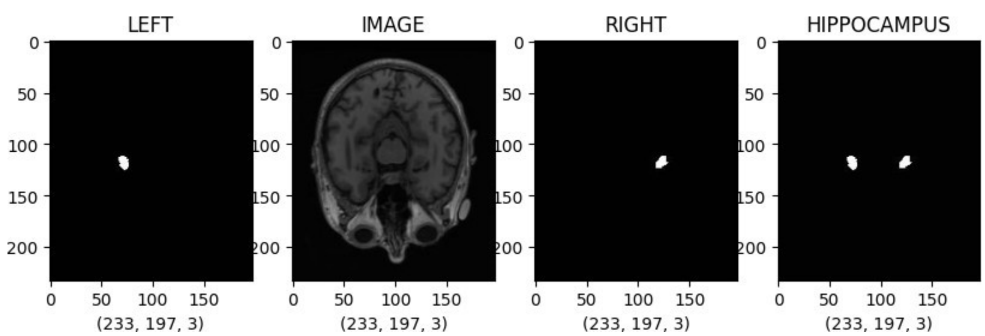

Figure 7: Hippocampus MRI dataset example

We preprocess the dataset through a custom class responsible for loading and modifying images and their corresponding masks. The images are transformed into tensors for further processing. The dataset is then split into training and validation sets using an 80:20 ratio. We create two DataLoader instances, one for each set. To assess our models' performance on different dataset sizes, we generate two variations: a smaller dataset with 300 randomly selected samples and batch size 10, and the entire dataset with all samples and batch size 64.

The data preparation process includes loading images and their hippocampus masks, applying a threshold, and ensuring sufficient pixels in the mask. Images and masks that meet the pixel count criteria are added to the dataset, extracting and preparing data for model training. Subsequently, images are padded to a target shape of 256 by default, ensuring consistent size for model input. Lastly, hippocampus masks are converted to binary masks using thresholding with a default value of 127. This results in a set of padded images and corresponding binary hippocampus masks, ready for training and validating the segmentation models.

II. U-Net/U-Net++ Architecture

The U-Net model consists of an encoder-decoder structure with skip connections between corresponding layers in the encoder and decoder. The encoder part, composed of convolutional blocks followed by max-pooling layers, extracts high-level semantic features from the input image. The decoder part, consisting of up-convolution layers and convolutional blocks, reconstructs the image at the original resolution while preserving spatial information. Skip connections help retain local information and improve the final segmentation by combining features from different levels of abstraction. The output layer uses a sigmoid activation function to produce pixel-wise binary segmentation masks.

The U-Net++ model improves upon the original U-Net by introducing nested skip pathways and deep supervision. It features an encoder-decoder structure with dense skip connections, which enhance information flow between different layers and levels of abstraction. The NestedConvBlock module consists of two convolutional layers, batch normalization, an activation function, and dropout. The UNetPlusPlus class defines the overall architecture, with multiple nested convolutional layers at each level. The deep supervision option allows the model to produce intermediate segmentation outputs at various depths, facilitating better gradient flow and improving the learning process. The final output is generated using a sigmoid activation function, producing pixel-wise binary segmentation masks.

III. Implement loss function

In our study, we employ Dice loss, Binary Cross-Entropy (BCE) loss, and a combination of both as our objective function. Dice loss is a similarity metric derived from the Dice coefficient, which ranges from 0 to 1, with 1 representing a perfect overlap between the prediction and ground truth. To mitigate potential division by zero errors, we introduce a smooth parameter within the Dice loss calculation. BCE loss, a well-known probabilistic loss function, computes the difference between predicted probabilities and true binary class labels. It is implemented using the BCELoss class and the binary cross-entropy with logits function from the PyTorch framework. Additionally, we define a combined loss function that leverages a weighted combination of Dice loss and BCE loss. This approach effectively addresses class imbalance issues and provides a robust training objective for our segmentation network. The weighting factors, alpha and beta, control the relative importance of the Dice loss and BCE loss, enabling us to fine-tune our model for optimal performance in diverse scenarios.

$$BCE = -(y \times log(p) + (1 - y) \times log(1 - p))$$

$$Dice\ Loss = 1 - \frac{(2 \times |A \cap B| + smooth)}{(|A| + |B| + smooth)}$$

$$Combined\ Loss = \alpha \times BCE + \beta \times DiceLoss$$

IV. Implement metrics

We utilize several performance metrics to evaluate our segmentation model, including Intersection over Union (IoU) score, Dice coefficient, numeric scores (false positive, false negative, true positive, and true negative), and accuracy score. To calculates the IoU score, a widely used metric to evaluate the accuracy of image segmentation models, as the ratio of the intersection to the union of the predicted output and the target. To computes the Dice coefficient, a similarity metric that measures the pixel-wise agreement between two binary images, as twice the intersection divided by the sum of pixels in both images. To assess the performance of binary classifiers, we calculate false positive (FP), false negative (FN), true positive (TP), and true negative (TN) scores. We also determine the accuracy score, a metric evaluating the overall performance of binary classifiers. The accuracy score is calculated as the ratio of the number of correct predictions to the total number of predictions made.

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{(TP + FP + FN)}$$

$$DSC = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} = \frac{2TP}{(2TP + FP + FN)}$$

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$$

V. Hyper Parameters

We experimented with various optimization algorithms and learning rate schedulers to train our U-Net++ model. For optimization, we tested the following setups: (1) Adam optimizer with a learning rate of 1e-4, (2) Stochastic Gradient Descent (SGD) with a learning rate of 1e-4 and a momentum of 0.99, and (3) Adam optimizer with a learning rate of 3e-4, betas of (0.9, 0.999), an epsilon of 1e-08, and a weight decay of 0. We also employed different learning rate schedulers, such as (a) Cosine Annealing Warm Restarts with an initial restart period T_0 of 10, a restart period multiplier T_mult of 2, and a minimum learning rate of 1e-6, (b) Cosine Annealing with a maximum period of 50 and a minimum learning rate of 1e-5, (c) Cosine Annealing with a maximum period of 1e10 and a minimum learning rate of 1e-5, and (d) Cyclic Learning Rate with a base learning rate of 1e-4, a maximum learning rate of 1e-2, a step size of 2000, and a 'triangular2' mode. By comparing the performance of the model under these different configurations, we aimed to identify the optimal hyper parameters for our segmentation task.

5 Result and Evaluation

In our experiments, we conducted an extensive evaluation of the U-Net and U-Net++ architectures for the task of Alzheimer's disease diagnosis using brain MRI images (Fig 8 & 9). Our results revealed that the U-Net++ model consistently outperformed the original U-Net architecture.

The superior performance of U-Net++ can be attributed to the incorporation of nested skip connections and deep supervision, which enables the model to capture more detailed feature representations and generalize better to previously unseen data. Although the U-Net++ model generally requires a longer training time compared to the original U-Net for the same dataset and number of epochs, this trade-off is justified by the improved segmentation results achieved by the U-Net++ architecture.

Our experiments indicate that the U-Net++ model is a promising tool for Alzheimer's disease diagnosis based on brain MRI segmentation. Its improved performance over the original

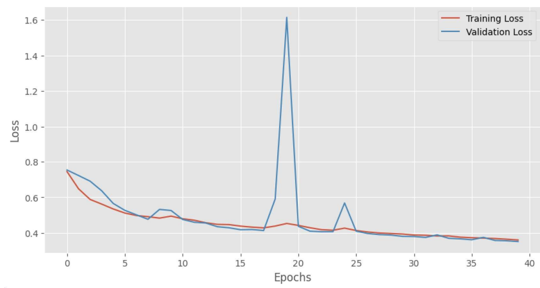

Figure 8: U-Net loss

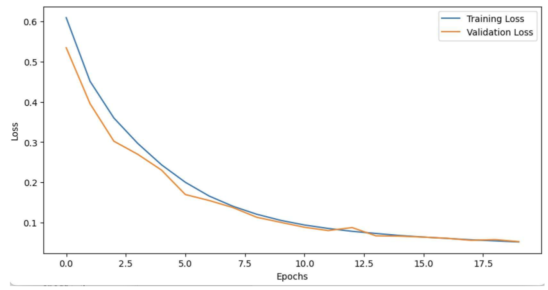

Figure 9: U-Net++ loss

U-Net highlights the effectiveness of nested skip connections and deep supervision in facilitating better feature extraction and generalization capabilities. Future work could explore additional architectural improvements, as well as alternative loss functions, to further enhance the performance of U-Net++ in this critical application.

To classifying Alzheimer's disease from brain MRI scans involves utilizing a dataset of tagged pictures to train a deep learning model. The InceptionV3 architecture, a pre-trained convolutional neural network (CNN) created by Google, is a popular option for this purpose. InceptionV3 perform exceptionally well in a variety of image classification tasks, including medical image analysis, because of its capacity to recognize intricate patterns and features in visual data. It is made up of a number of inception modules that allow the model the ability to pick up several layers of abstraction at once, enabling it to effectively manage a range of scales and complexities in the data input.

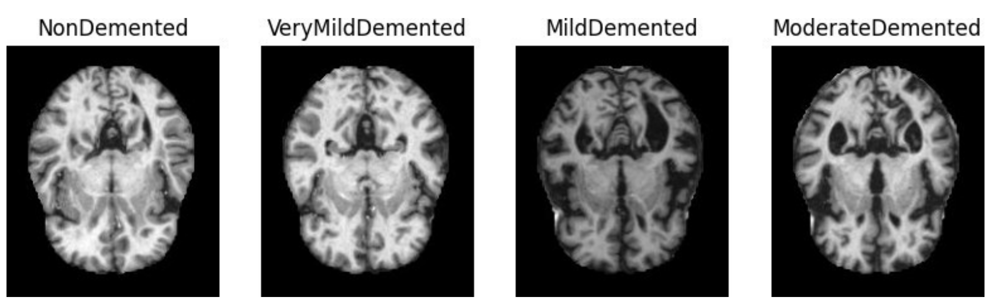

Figure 10: Four classes of MRI pictures

The InceptionV3 model can trained with a dataset obtained from Kaggle that has four classes of MRI pictures (Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented) for the classification of AD (Fig 10). By comparing the projected class labels with the actual class labels from the testing set, the classification accuracy can be assessed. Custom callback functions and learning rate reduction techniques, such ReduceLROnPlateau, can be used to improve the training process by stabilizing the training and terminating it when the target accuracy is attained (Fig 11).

We achieved outstanding performance metrics across all four classes of AD using the InceptionV3 architecture and unique callback functions in our deep learning model (Fig 12). Overall performance indicators show that our model routinely outperforms the competition in the categorization job.

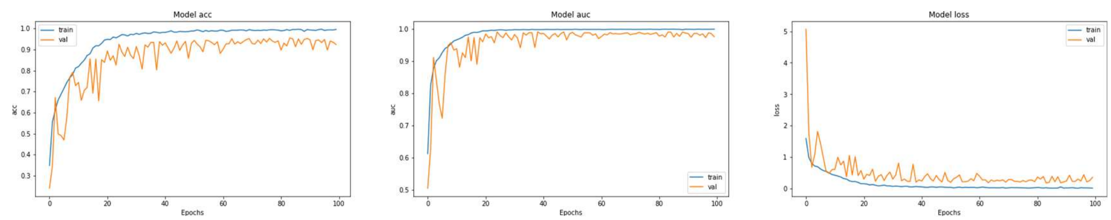

Figure 11: InceptionV3 Loss

	precision	recall	f1-score	support
NonDemented	0.86	1.00	0.92	639
VeryMildDemented	1.00	1.00	1.00	635
MildDemented	0.91	0.88	0.89	662
ModerateDemented	0.94	0.82	0.88	624
micro avg	0.92	0.92	0.92	2560
macro avg	0.93	0.92	0.92	2560
weighted avg	0.93	0.92	0.92	2560
samples avg	0.92	0.92	0.92	2560

Figure 12: InceptionV3 Metrics

6 Future Work

We intend to further our inquiry towards the diagnosis of Alzheimer's disease (AD) in next work employing sophisticated U-Net variations trained on brain MRI scans. R2U-Net [10], Attention U-Net [11], Trans U-Net [12], and 3D U-Net [13] are a few interesting designs that have shown improved performance in a variety of medical picture segmentation tasks. We anticipate that combining these cutting-edge models will further increase the precision and robustness of AD diagnosis.

In order to capture both local and global contextual information in the input images, R2U-Net (Fig 13) first combines the U-Net framework with recurrent convolutional layers. We predict that adding these recurrent layers will improve the segmentation of brain anomalies associated to AD.

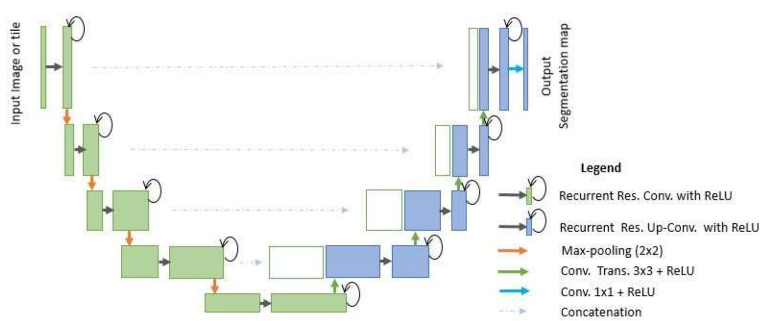

Figure 13: R2U-Net

Similar to this, Attention U-Net (Fig 14) modifies the U-Net design to include an attention mechanism that enables the model to concentrate on pertinent regions during segmentation. This attention mechanism will be modified for our goal in an effort to improve the model's capacity to recognize and define important AD-related traits.

Trans U-Net (Fig 15) also makes use of the U-Net framework's transformers, which have the ability to record far-reaching dependencies and intricate interactions between picture regions. Given the intricate spatial patterns connected to AD, we think that adding transformers to

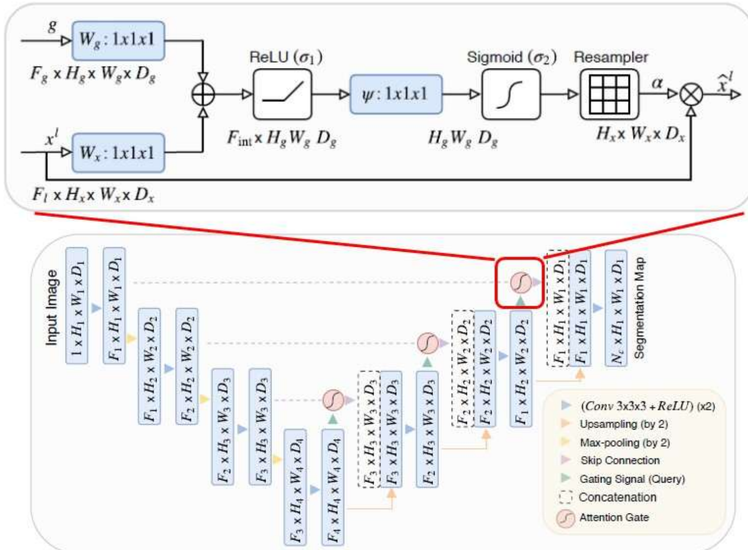

Figure 14: Attention U-Net

our model can enhance diagnostic performance.

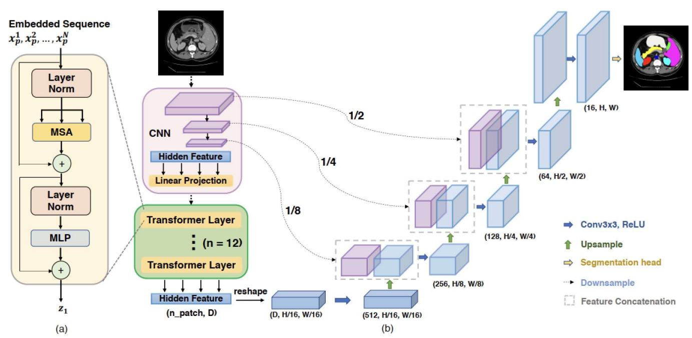

Figure 15: Trans U-Net

Additionally, we intend to look into the use of 3D U-Net (Fig 16), which processes volumetric data, to more accurately capture the three-dimensional structure of brain MRI scans, possibly resulting in a more precise AD diagnosis.

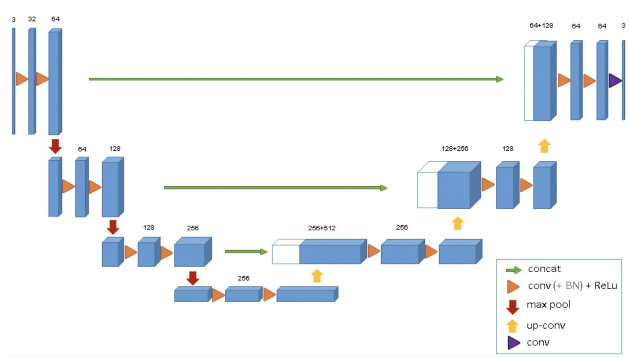

Figure 16: 3D U-Net

In the future, we want to work with researchers and physicians to confirm how well our models work in actual environments. Through the use of an interdisciplinary approach, we can guarantee that our models generalize well to a wide range of patient groups and imaging techniques, ultimately improving their therapeutic utility. Long-term, we want to investigate how our models may be used in the early diagnosis, planning, and monitoring of various neurological illnesses. We intend to contribute to the creation of cutting-edge AI-driven tools that will change the detection and treatment of Alzheimer's disease and other neurological illnesses by building on the groundwork established in this study.

7 Reference

[1] Goedert M, Spillantini MG: A century of Alzheimer's disease. Science. 2006, 314: 777-781. 10.1126/science.1132814.

[2] Bisaz R, Travaglia A, Alberini CM. The neurobiological bases of memory formation: from physiological conditions to psychopathology. Psychopathology. 2014;47(6):347–356. doi:10.1159/000363702

[3] Preston AR, Eichenbaum H. Interplay of hippocampus and prefrontal cortex in memory. Curr Biol. 2013;23(17):R764–R773. doi:10.1016/j.cub.2013.05.041

[4] Jahn H. Memory loss in Alzheimer's disease. Dialogues Clin Neurosci. PMID: 24459411

[5] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015.

[6] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh and J. Liang, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," in IEEE Transactions on Medical Imaging, vol. 39, no. 6, pp. 1856-1867, June 2020, doi: 10.1109/TMI.2019.2959609. **paper github code**

[7] Hane, Francis T., Brenda Y. Lee, and Zoya Leonenko. "Recent progress in Alzheimer's disease research, part 1: Pathology." Journal of Alzheimer's Disease 57.1 (2017): 1-28.

[8] Mu, Yangling, and Fred H. Gage. "Adult hippocampal neurogenesis and its role in Alzheimer's disease." Molecular neurodegeneration 6.1 (2011): 1-9.

[9] Meyer, Michael G., et al. "The Cell-CT 3-dimensional cell imaging technology platform enables the detection of lung cancer using the noninvasive LuCED sputum test." Cancer cytopathology 123.9 (2015): 512-523.

[10] Alom, Md Zahangir, et al. "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation." arXiv preprint arXiv:1802.06955 (2018).

[11] Oktay, Ozan, et al. "Attention u-net: Learning where to look for the pancreas." arXiv preprint arXiv:1804.03999 (2018).

[12] Guo, Shirong, et al. "Trans-U: Transformer Enhanced U-Net for Medical Image Segmentation." 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA). IEEE, 2022.

[13] Ç i ¸cek, Ö zgü n, et al. "3D U-Net: learning dense volumetric segmentation from sparse annotation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. Springer International Publishing, 2016.