UNet++ with Attention Mechanism for Hippocampus Segmentation

Xueting Cui
Electrical and Computer Engineering
Florida International University
Miami FL, USA
xcui005@fiu.edu

Thony Yan Liang
Electrical and Computer Engineering
Florida International University
Miami FL, USA
tyan001@fiu.edu

Marymossadat Aghili
Microsoft
San Francisco Bay Area
California, USA
maryam.aghili85@gmail.com

Micheal Adeyosoye
Electrical and Computer Engineering
Florida International University
Miami FL, USA
madeyo004@fiu.edu

Rosie E. Curiel Cid
Department of Neuropsychology
University of Miami
Miami FL, USA
rcuriel2@med.miami.edu

David Lowenstein
Center for Cognitive Neuroscience and Aging
University of Miami
Miami FL, USA
dloewenstein@med.miami.edu

Ranjan Duara
Wien Center for Alzheimer's Disease and Memory Disorders
Mt Sinai Medical Center
Miami FL, USA
ranjan.duara@msmc.com

Malek Adjouadi
Electrical and Computer Engineering
Florida International University
Miami FL, USA
adjouadi@fiu.edu

***Abstract*—Analyzing the hippocampus in the brain through magnetic resonance imaging (MRI) plays a crucial role in diagnosing and making treatment decisions for several neurological diseases. Hippocampus atrophy is among the most informative early diagnostic biomarkers of Alzheimer's disease (AD), yet its automatic segmentation is extremely difficult given the anatomical structure of the brain and the lack of any contrast in between its different regions. The gold standard remains manual segmentation and the use of brain atlases. In this study, we use a well-known image segmentation model, UNet++, and introduce an attention mechanism called the Convolutional Block Attention Module (CBAM) to the UNet++ model. This integrated model improves the feature weights of our region of interest, and hence increases the accuracy in segmenting the hippocampus. Results show averages of 0.8715, 0.8107, 0.8872, and 0.9039 for the metrics of Dice, Jaccard, Precision, and Recall, respectively.**

***Index Terms*—Hippocampus segmentation, Alzheimer's disease, MRI, UNet++, Attention Mechanism**

I. Introduction

As society ages, Alzheimer's disease (AD) will affect more people and families. Projections indicate that by 2050, more than 13.8 million people in the United States will have dementia[1]. In addition, AD is irreversible and can cause severe memory and behavioral problems.Therefore, early detection of AD and its initial stage, namely mild cognitive impairment (MCI), and effective diagnosis and treatment planning, especially using computer-assisted methods, have attracted a lot of attention in recent years[2]. The studies[3][4] showed that hippocampal volume is an essential quantitative indicator that may be used as a biomarker for neurological diseases such as Alzheimer's disease. Therefore, correct segmentation of this region is critical to assess any subtle structural and volumetric changes in the brain to diagnose early this disease.

In the medical field, doctors use MRI to describe the internal structure and state of the brain to analyze brain diseases. At present, artificial intelligence technology plays an increasingly significant role in the field of medical image processing. Therefore, creating a robust deep learning model for image segmentation with great accuracy and efficiency will significantly impact the treatment of Alzheimer's disease. Among them, UNet[5] performs well in image segmentation, and UNet++[6] performs better as an improved version.

Our method introduces an attention mechanism in the UNet++ network, and the model obtained after training achieves satisfactory performance on the hippocampus segmentation task in MRI in comparison to Unet and Unet++ without the attention mechanism.

II. Methodology

In this section, the methodology proposed for the neural network architecture is expressed visually through Fig. 1. What follows is a retrospective on the UNet++ with attention mechanism (CBAM) and the loss function used for our experiments.

A. UNET++ with CBAM

Based on the traditional UNet encoding and decoding U-shaped structure, UNet++ introduces a series of convolutional blocks that are used to bridge the semantic gap between the corresponding feature maps of encoding and decoding

2769-5654/22/$31.00 ©2022 IEEE
DOI 10.1109/CSCI58124.2022.00312

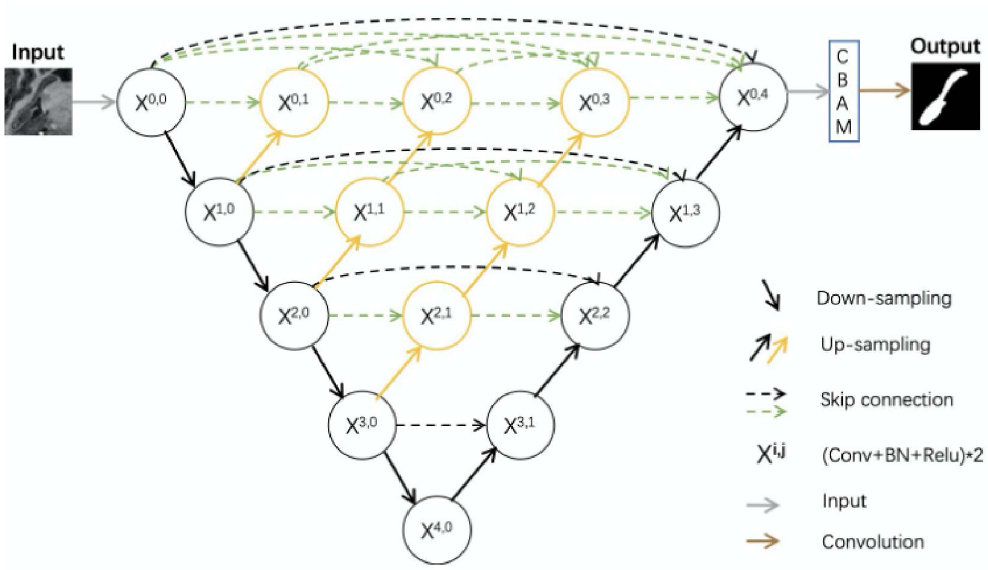

Fig. 1. Network structure of our UNet++ implementation with CBAM

in UNet before feature fusion. For example, in Fig. 1, the semantic difference between $X^{0,0}$ and $X^{1,3}$ is filled by the convolutional blocks between them. The black nodes and lines in Fig. 1 represent the original UNet network. Its structure mainly includes the encoding part and the decoding part. First, the encoding part is formed by the arrows pointing downward starting from $X^{0,0}$ on the left. In this part, $X^{4,0}$ is obtained by a series of down-sampling from $X^{0,0}$. Then starting from $X^{4,0}$, the decoding part formed by the upward arrows on the right is completed after a series of up-sampling operations.

The inner yellow and green parts are the added parts that make up UNet++. For example, the yellow node $X^{2,1}$ is obtained from $X^{2,0}$ and $X^{3,0}$. Node $X^{3,0}$ after up-sampling is concatenated with $X^{2,0}$ to serve as input data of $X^{2,1}$. Similarly for $X^{0,0}$ and $X^{1,3}$, with $X^{1,3}$ after up-sampling being concatenated with $X^{0,0}$ to serve as input data of $X^{0,4}$ in UNet. But the features of $X^{0,0}$ have a semantic gap with the features of $X^{1,3}$ after going through the U shape. UNet++ introduces a series of intermediate nodes between $X^{0,0}$ and $X^{1,3}$ to make up for this difference. In UNet++, $X^{1,3}$ after up-sampling is concatenated with $X^{0,0}$, $X^{0,1}$, $X^{0,2}$ and $X^{0,3}$ as the input data of $X^{0,4}$. Therefore, the semantic gap between the feature maps of $X^{0,0}$ and $X^{1,3}$ is bridged.

We introduce the attention mechanism model CBAM, Convolutional Block Attention Module[7], to enhance the feature map and improve semantic segmentation accuracy. The CBAM model enables the network to quickly locate the ROI in the feature map, which is then analyzed in detail. Due to the small size of our images, we only introduced CBAM before the final convolution operation in decoding. First, the feature map generated by $X^{0,4}$ is processed by CBAM. Then a convolution operation is performed to obtain the output result.

B. Loss Functions

The loss function used for our architecture is the BCEDiceLoss function. This loss function is a combination of the Binary Cross Entropy (BCE) and the Dice Loss function. This combination overcomes the shortcomings that comes with both loss functions when used separately, so as to consolidate the benefits of the class imbalance (to deal with imbalanced class distributions) from the Dice Loss function and smooth curve (avoiding discontinuities and outliers) from the BCE function.

$$BCEDiceLoss = BCE + DiceLoss \tag{1}$$

The BCE loss function determines the difference between two probability distributions. The BCE can be very useful for classification tasks that are binary in nature (i.e., two choices). In this particular endeavor, image segmentation is pixel-level classification, with 1s and 0s as outcomes tomean object and background or vice versa. So this loss function would work well for the proposed neural network architecture. To deal with these 0's and 1's outcomes or outputs, the Binary Cross-Entropy is defined as follows:

$$L_{BCE}(y, \widehat{y}) = \ -(ylog(\widehat{y}) + (1-y)log(1-\widehat{y})) \tag{2}$$

where y represents the true value, and $\widehat{y}$ represents the predicted outcome.

The dice coefficient is a reproducible validation metric that looks at the spatial overlap index. It indicates whether two segmented results are overlapping or not using a range from 0 to 1. Eventually, it was implemented as a loss function called the Dice Loss function[8] as defined below:

$$DL(y, \widehat{y}) = 1 - \frac{2y\widehat{y} + 1}{y + \widehat{y} + 1} \tag{3}$$

In this function, one is added in both the numerator and the denominator of the fraction part so we can make sure that the function is defined in the edge case scenarios, for example, when $y = \hat{p} = 0$[9].

III. Experiments & Results

A. Dataset

In this study, the data used for training our segmentation model were obtained from the Medical Segmentation Decathlon Challenge[10]. The MRI provided by the dataset is T1-weighted. The ROIs were subdivided into the anterior and posterior parts of the hippocampus. Therefore, in the label image corresponding to MRI, there are two types of label values. In our experiments, we processed the label data by merging the anterior and posterior parts, considering them as the entire hippocampus. There were 260 brain MRIs available for this study. They were divided into 156(60%), 52 (20%), and 52(20%) MRIs for training, validation, and testing, respectively.

B. Evaluation Metrics

To quantitatively evaluate the proposed method and compare the performance with others, we used four standard metrics in this study. First, we use the Mean Dice Similarity Coefficient (DSC) to measure overlaps between the ground truth label A_g and the predicted label A_p.

$$DSC = 2 \times \sum_{i=1}^{n} \frac{|A_{p_i} \cap A_{g_i}|}{|A_{p_i}| + |A_{g_i}|} \tag{4}$$

Jaccard Similarity Coefficient (JSC) is also called intersection over union (IoU) coefficient, that is the intersection of A_g and A_p, divided by their union as expressed in (5) yields a similarity measure between A_g and A_p.

$$JSC = \sum_{i=1}^{n} \frac{|A_{p_i} \cap A_{g_i}|}{|A_{p_i}| + |A_{g_i}| - |A_{p_i} \cap A_{g_i}|} \tag{5}$$

Precision Index is the intersection between A_g and A_p over ground truth label A_g.

$$Precision\ Index = \sum_{i=1}^{n} \frac{|A_{p_i} \cap A_{g_i}|}{|A_{g_i}|} \tag{6}$$

Recall Index is the intersection between A_g and A_p labels over predicted label A_p.

$$Recall\ Index = \sum_{i=1}^{n} \frac{|A_{p_i} \cap A_{g_i}|}{|A_{p_i}|} \tag{7}$$

C. Results and Analysis

We conducted three different experiments to assess the model's performance. In each experiment, we trained the model three times to obtain an average of the metric data. All architecture models were trained with the same dataset. In the first experiment, we used the base architecture model UNet without the CBAM. Next, we used UNet++ without the CBAM. Finally, we used both the UNet++ and the CBAM. The four metrics of DSC, JSC, Precision Index, and Recall Index were used to evaluate the performance of each architecture. The average evaluation scores of the three networks are shown in Table 1. We can see that the values of the four evaluation indicators of the proposed method are all better than the other two network models. The comparisons of the segmentation results among the three models are shown in Fig. 2. The images in Fig. 2 are from the test set, showing examples of two different patients for visual appreciation.

The results of the experiments show that the proposed method which integrates Unet++ with attention mechanism achieves the best results in comparison when Unet and Unet++ are used alone. From the results shown in Table 1, the proposed integrated method outperforms the other two networks on average on the brain MRI dataset.

TABLE I
Metrics results

	CBAM	Exper.	Dice	Jaccard	Prec.	Recall
UNet	No	1	0.8687	0.8004	0.8822	0.8962
		2	0.8703	0.8017	0.8819	0.8982
		3	0.8668	0.8028	0.8812	0.9003
		avg.	**0.8686**	**0.8016**	**0.8818**	**0.8982**
UNet++	No	1	0.8708	0.8042	0.8779	0.9054
		2	0.8716	0.8050	0.8852	0.8989
		3	0.8666	0.8054	0.8868	0.8976
		avg.	**0.8697**	**0.8049**	**0.8833**	**0.9006**
UNet++ Proposed Method	Yes	1	0.8714	0.8105	0.8849	0.9061
		2	0.8699	0.8108	0.8885	0.9025
		3	0.8731	0.8109	0.8883	0.9030
		avg.	**0.8715**	**0.8107**	**0.8872**	**0.9039**

IV. Conclusion

Our goal in this study is to automatically segment the hippocampus using a labeled MRI dataset. The deep learning method as proposed is based on UNet++ with the inclusion of an attention mechanism. We conducted experiments on the brain MRI dataset for hippocampus segmentation. We achieved better performance compared to the two previous networks for image segmentation, demonstrating the effectiveness of our proposed method. Furthermore, our experimental results show that the attention mechanism has had an impact on the Convolutional Neural Network, improving the performance of processing image pixel details. As with the general performance of CNNs for various image processing, we predict that in addition to brain MRI, other imaging modalities such as arterial spin labelling (ASL), diffusion tensor imaging (DTI), HighResHippocampus, susceptibility weighted imaging (SWI), and T2Flair, to mention a few, could be co-registered to the T1 weighted MRIs to enhance the prospects for a more accurate automated segmentation that could be more contextual through a multimodal imaging process.

Moreover, to better appreciate the results summarized in Table 1, the study reported in [11] provides all the evidence we need in the challenge faced when segmenting brain regions in MR images. Interestingly, the authors of this study show that even when there is very high agreement among four expert tracers (pairwise Jaccard indices 0.82-0.87), the volumetric

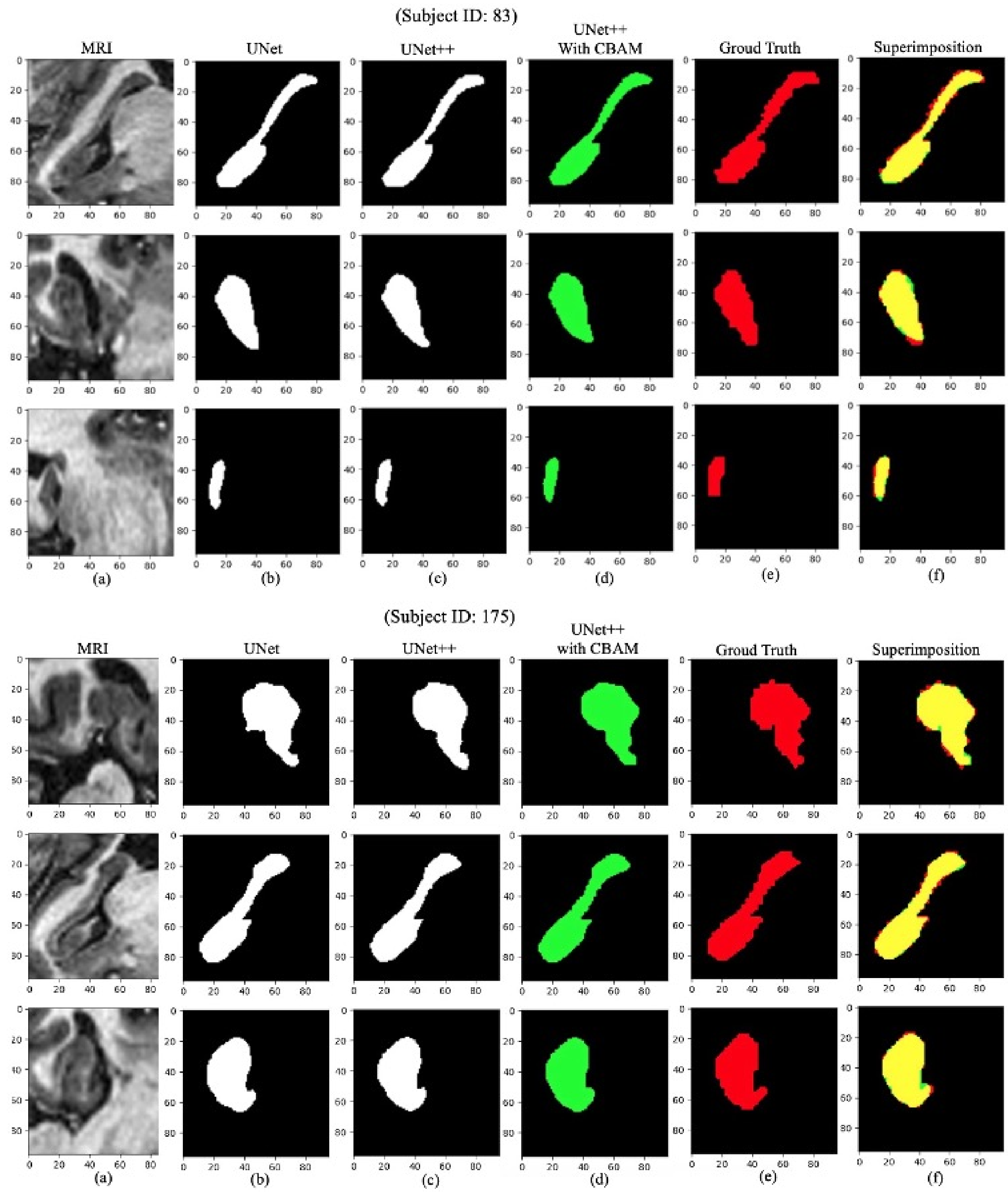

Fig. 2. (a) Input MRI. (b) Result of UNet segmentation. (c) Result of UNet++ segmentation. (d) Segmentation result of our method. (e) Ground truth. (f) Superimposition of the ground truth and our model's segmentation. The yellow represents the pixels where our model matched correctly the ground truth.

results among the four expert tracers obtained using the HarP (Hierarchical Attention Oriented Region-Based Processing) benchmark dataset consisting of 135 MRIs still showed a mean volume difference of 9% between them with a standard deviation of 7%.

Acknowledgments

This research is supported by the National Science Foundation under grants: CNS-1920182, CNS-2018611, and CNS-1551221, and with the National Institutes of Health - National Institute on Aging (NIA) through the P30AG066506 with the 1Florida Alzheimer's Disease Research Center (ADRC).

References

[1] Alzheimer's Association. 2019 alzheimer's disease facts and figures. *Alzheimer's & dementia*, 15(3):321–387, 2019.

[2] Mehdi Shojaie, Solale Tabarestani, Mercedes Cabrerizo, Steven T DeKosky, David E Vaillancourt, David Loewenstein, Ranjan Duara, and Malek Adjouadi. Pet imaging of tau pathology and amyloid-, and mri for alzheimer's disease feature fusion and multimodal classification. *Journal of Alzheimer's disease : JAD*, 84(4):1497—1514, 2021.

[3] Hongming Li, Mohamad Habes, David A Wolk, Yong Fan, and Alzheimer's Disease Neuroimaging Initiative and the Australian Imaging Biomarkers and Lifestyle Study of Aging. A deep learning model for early prediction of alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's amp; dementia : the journal of the Alzheimer's Association*, 15(8):1059—1070, August 2019.

[4] Davis C Woodworth, Hannah L Nguyen, Zainab Khan, Claudia H Kawas, María M Corrada, and S Ahmad Sajjadi. Utility of mri in the identification of hippocampal sclerosis of aging. *Alzheimer's amp; dementia : the journal of the Alzheimer's Association*, 17(5):847—855, May 2021.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[6] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018.

[7] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[8] Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sébastien Ourselin, and M. Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *CoRR*, abs/1707.03237, 2017.

[9] Shruti Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, oct 2020.

[10] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, AnnetteKopp-Schneider, Bennett A. Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M. Summers, Bram van Ginneken, Michel Bilello, Patrick Bilic, Patrick F. Christ, Richard K. G. Do, Marc J. Gollub, Stephan H. Heckers, Henkjan Huisman, William R. Jarnagin, Maureen K. McHugo, Sandy Napel, Jennifer S. Goli Pernicka, Kawal Rhode, Catalina Tobon-Gomez, Eugene Vorontsov, Henkjan Huisman, James A. Meakin, Sebastien Ourselin, Manuel Wiesenfarth, Pablo Arbelaez, Byeonguk Bae, Sihong Chen, Laura Daza, Jianjiang Feng, Baochun He, Fabian Isensee, Yuanfeng Ji, Fucang Jia, Namkug Kim, Ildoo Kim, Dorit Merhof, Akshay Pai, Beomhee Park, Mathias Perslev, Ramin Rezaiifar, Oliver Rippel, Ignacio Sarasua, Wei Shen, Jaemin Son, Christian Wachinger, Liansheng Wang, Yan Wang, Yingda Xia, Daguang Xu, Zhanwei Xu, Yefeng Zheng, Amber L. Simpson, Lena Maier-Hein, and M. Jorge Cardoso. The medical segmentation decathlon, 2021.

[11] Marina Boccardi, Martina Bocchetta, Rossana Ganzola, Nicolas Robitaille, Alberto Redolfi, Simon Duchesne, Clifford R Jack Jr, Giovanni B Frisoni, EADC-ADNI Working Group on The Harmonized Protocol for Manual Hippocampal Segmentation, for the Alzheimer's Disease Neuroimaging Initiative, George Bartzokis, et al. Operationalizing protocol differences for eadc-adni manual hippocampal segmentation. *Alzheimer's & Dementia*, 11(2):184–194, 2015.