



# 3D HIPPOCAMPUS SEGMENTATION USING A HOG BASED LOSS FUNCTION WITH MAJORITY POOLING

Arijit De<sup>1</sup> , Mona Tiwari<sup>2</sup>, Ananda S. Chowdhury<sup>1</sup> 

<sup>1</sup>Department of Electronics and Telecommunication Engineering, Jadavpur University  
Kolkata, West Bengal, India

<sup>2</sup>Department of Neuro-radiology, Institute of Neurosciences  
Kolkata, West Bengal, India

## ABSTRACT

Hippocampus (HC) segmentation plays a key role in diagnosis of predominant neuro-degenerative diseases like Alzheimer's, Parkinson's and common neurological disorders like Epilepsy. In this paper, we propose a solution to the 3D HC segmentation problem from the MRI data using shape driven loss function and attention Unet. In particular, a Histogram of Oriented Gradients (HOG) based formulation is developed to extract shape features. We suggest a pooling technique as a substitute to the histogram calculation for HOG. This is to address the problem that histogram is not derivable thereby making the error in a loss function from histogram unsuitable for back propagation in deep learning models. The performance of our proposed model is validated on two publicly available datasets, namely, HarP and Kulaga-Yoskovitz (KY). Our segmentation accuracy with a dice similarity score of 0.947 and 0.923 in HarP and KY respectively is found to outperform the attention UNet model with only Dice loss, and, a number of state-of-the-art approaches.

**Index Terms**— Hippocampus Segmentation, HOG features, Attention Unet, Majority Pooling, Loss function.

## 1 Introduction

The hippocampus (HC) is a bilateral brain structure located in the medial temporal lobe at both sides of the brainstem near the cerebellum. HC is a very intricate and heterogeneous structure broadly divided into several sub-fields, namely, subiculum, cornu ammonis (CA1/2/3/4), and dentate gyrus (DG). Alzheimer's Disease (AD), Parkinson's Disease (PD), autism, multiple sclerosis, and natural ageing affect the HC. Stages of AD affect hippocampal sub-fields differently. The atrophy of stratum radiatum/stratum lacunosum-moleculare (SRLM) apical dendrites of hippocampal CA1 closely mirrors episodic memory impairment in AD as well as PD patients [1]. HC atrophy study may give biomarkers and improved techniques for detecting and predicting AD and other neurodegenerative illnesses. So, segmentation of HC assumes paramount importance in neuroradiology.

The result of manual segmentation is usually regarded as the gold standard because of its high accuracy. But, such a strategy becomes extremely time-consuming and laborious. This high expense of manual segmentation has prompted the development of efficient automation techniques. FreeSurfer (FSL) [2] was one of the early attempts. FSL is capable of producing decent coarse segmentation. However, it lacks finer details and accurate edge delineation. Some other noteworthy works attempted HC segmentation using atlas registration [3] and level sets [4], but they were in 2D and did not perform well in 3D. In recent times, several deep learning (DL) techniques have been used to improve HC segmentation. In recent times, Convolutional Neural Networks (CNN) gained popularity over previous semi-automatic and automated approaches for HC segmentation [5].

Medical imaging lacks consistent and sufficient annotated data, making DL algorithms performing sometimes below par [6]. To circumvent this limitation, shape-driven DL algorithms have evolved, which can combine CNN-learned deep features with structural shape information, to boost the performance [7]. Only a few shape-based HC segmentation studies are reported till date. Brusini et al. [8] employed shape fitting and UNet to segment the HC. However, they used three 2D models instead of one 3D model, which requires more training time. Others like Tang et al. [9] have employed signed distance maps (SDM) to build loss functions that penalise Dice loss and SDM values. The above technique estimated loss by training a model to predict SDM and calculating dice loss between the predicted and the ground truth SDM. This required additional model training.

In this paper, we propose a novel shape-based DL strategy for HC segmentation. Contour directional information is incorporated to create a shape based model using Histogram of Oriented Gradients (HOG). HOG [10] is an excellent feature descriptor that can extract shape as well as structural information from an image by computing magnitude and angle gradients. The application of using HOG with deep learning is well demonstrated in [11]. We took inspiration from [12] while developing the formalism. We have further suggested an effi-

cient pooling to tackle the derivability problem of histogram in HOG. Normal pooling methods like max-pool, min-pool and average-pool may not yield desirable results in this situation. Our main contributions are summarized below -

1. We have incorporated shape based loss using 3D gradient magnitude and gradient angles of HoG to find the 3D HC boundary along with directional information.
2. A 3D weighted majority pooling technique, that can act as a substitute for histogram operation on tensor data, is developed. This can be implemented in DL models as it is derivable.
3. Our proposed solution for HC segmentation can provide highly accurate results in 3D. This can reduce the diagnosis time and expedite the surgical planning.

## 2 Proposed Method

In this section, we describe our proposed method. The block diagram showing the overall method is presented in Fig. 1.

### 2.1 Attention UNet

3D Unet [13], a deep fully convolutional encoder-decoder network architecture has gained popularity for voxel-wise segmentation. In this work, we utilise the Attention based UNet as described in [14]. The authors in [15] have argued that it is one of the best Unet models in terms of accuracy, number of parameters, training and inference time. The attention mechanism is necessary for training our model with a focus on the HC region. This focus has to be explicitly induced due to the inherent difficulty in distinguishing HC from its neighboring tissues, which exhibit similar contrast.

Our Attention Unet model has a depth of 4 and takes as input a 3D single channel tensor of dimension  $(1 \times \text{height} \times \text{width} \times \text{depth})$ . The model outputs a 3D single channel tensor of same dimensions containing the segmentation mask. We use batch normalization for each convolutional layer followed by ReLU activation.

### 2.2 3D HOG Features

HOG is a well-known feature descriptor which is able to extract structural and shape information [10]. It is better than other edge descriptors as it uses magnitude as well as angle of the gradient to compute the features. Since, we deal with 3D volumetric images, we use 3D feature descriptors which are shown to carry more information than their 2D counterparts [16].

We first extract gradients from an image volume  $A$  of dimension  $h \times w \times d$ ; where  $h$ ,  $w$  and  $d$  respectively denote the height, width and depth of  $A$ . We calculate  $G_x$ ,  $G_y$  and  $G_z$ , the gradient matrices along the x, y and z dimensions respectively by convolving  $A$  with a block of  $3 \times 3 \times 3$  kernel. The elements of the matrices are as follows:

$$\begin{aligned} G_x(i, j, k) &= I(i+1, j, k) - I(i-1, j, k) \\ G_y(i, j, k) &= I(i, j+1, k) - I(i, j-1, k) \\ G_z(i, j, k) &= I(i, j, k+1) - I(i, j, k-1) \end{aligned} \quad (1)$$

Here,  $i, j, k$  denote coordinates along the  $x, y, z$  directions.  $I(i, j, k)$  is the intensity at  $A(i, j, k)$ . We build three matrices denoting the magnitude ( $V_\mu$ ), the azimuth angle ( $V_\theta$ ) and the zenith angle ( $V_\phi$ ) of the gradient. The elements of the matrices are respectively given by:

$$V_\mu(i, j, k) = \sqrt{G_x^2 + G_y^2 + G_z^2} \quad (2)$$

$$V_\theta(i, j, k) = \left| \tan^{-1} \frac{G_y}{G_x} \right| \quad (3)$$

$$V_\phi(i, j, k) = \left| \cos^{-1} \frac{G_z}{V_\mu} \right| \quad (4)$$

Naturally, each of  $V_\mu, V_\theta, V_\phi$  has the dimension of  $A$ , i.e.,  $(h \times w \times d)$ .

### 2.3 Majority Pooling

The histogram, which is an integral part of HOG, is not a derivable function. This is a critical problem as for deep learning models, we need derivable functions to back propagate the loss. Hence, we present a novel majority pooling technique that pools the magnitude of the voxels with the most frequently occurring angle (separately in  $V_\theta$  &  $V_\phi$ ).

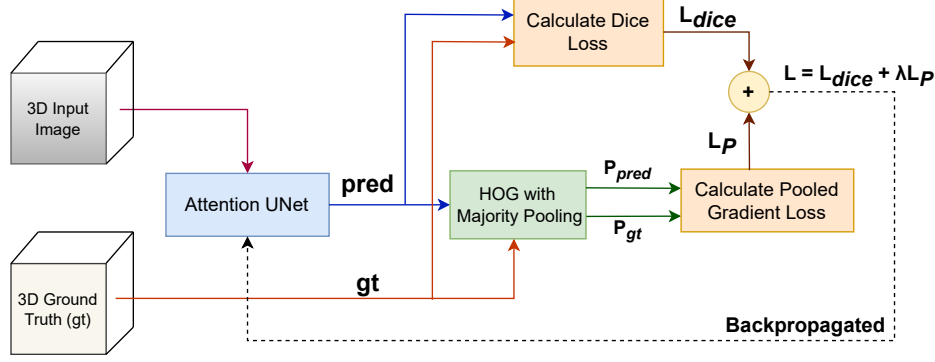
We first divide the three matrices  $V_\mu, V_\theta$  &  $V_\phi$  into  $n \times n \times n$  voxel grids (smaller matrices). So, each grid will contain  $\frac{h}{n} \times \frac{w}{n} \times \frac{d}{n}$  voxels. Let us consider the total number of such grids as  $g$ . Then, we can write  $V_\mu = [v_\mu^1, v_\mu^2, \dots, v_\mu^g]$  where the smaller matrices  $v_\mu^1, v_\mu^2, \dots, v_\mu^g$  are constructed row-wise with a stride of  $n$  and each of them has dimensions  $n \times n \times n$ . Similarly, we write  $V_\theta = [v_\theta^1, v_\theta^2, \dots, v_\theta^g]$  and  $V_\phi = [v_\phi^1, v_\phi^2, \dots, v_\phi^g]$ . We next construct two sets  $\psi_\theta$  and  $\psi_\phi$  to respectively select the voxels from  $v_\theta^s$  and  $v_\phi^s$  that have non-zero magnitudes in the corresponding location of  $v_\mu^s$  ( $s = 1, 2, \dots, g$ ). So, we write:

$$\begin{aligned} \psi_\theta &= \{v_\theta^s(i, j, k) | v_\mu^s(i, j, k) > 0\} \\ \psi_\phi &= \{v_\phi^s(i, j, k) | v_\mu^s(i, j, k) > 0\} \end{aligned} \quad (5)$$

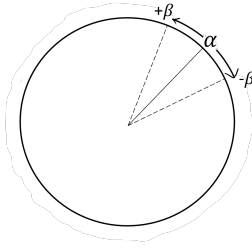
where,  $v_\theta^s(i, j, k)$ ,  $v_\phi^s(i, j, k)$  and  $v_\mu^s(i, j, k)$  is the value at location  $(i, j, k)$  of these matrices ( $i = 1 \dots n, j = 1 \dots n, k = 1 \dots n$ ). We now find the most frequently occurring angle  $\alpha$  from  $\psi_\theta$  and save the voxel location of all the angles in the range  $[\alpha - \beta, \alpha + \beta]$  (see Fig. 2) into  $M_\theta$ , where  $M_\theta \subset \psi_\theta$ . This is done to pool magnitude values from  $v_\mu^s$  with corresponding angles in  $v_\theta^s$  that are equal to or close to  $\alpha$ . This range is suitable as the gradient directions of all the angles within this range give a similar idea about the main direction of that region. We construct  $M_\phi$  likewise, where,  $M_\phi \subset \psi_\phi$ .

Finally, we create matrices  $p_\theta^s, p_\phi^s$  to pool only those values from  $v_\mu^s$  contained in  $M_\theta$  and  $M_\phi$  respectively and set the values at other locations to zero. So, we write:

$$\begin{aligned} p_\theta^s(i, j, k) &= \begin{cases} v_\mu^s(i, j, k) & \text{if } (i, j, k) \in M_\theta \\ 0 & \text{otherwise} \end{cases} \\ p_\phi^s(i, j, k) &= \begin{cases} v_\mu^s(i, j, k) & \text{if } (i, j, k) \in M_\phi \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (6)$$



**Fig. 1:** Block diagram of our proposed approach:  $pred$  denotes the predicted output of Attention Unet model and  $gt$  denotes the ground truth.  $P_{pred}$  and  $P_{gt}$  are the pooled gradients of predicted and ground truth respectively, as shown in Eq. 8.



**Fig. 2:** Figure showing the range of angles considered for selecting the voxels from  $v_{\theta}^s$  and  $v_{\phi}^s$ . If  $\alpha$  is the most frequently occurring angle, then all angles in range  $\alpha \pm \beta$  are considered.

where  $p_{\theta}^s(i, j, k)$  is the value at location  $(i, j, k)$  of  $p_{\theta}^s$  and similarly for  $p_{\phi}^s$ . All  $p_{\theta}^s$  ( $s = 1 \dots g$ ) of dimension  $n \times n \times n$  are concatenated row-wise to form the pooling matrix  $P_{\theta}$  of dimension  $h \times w \times d$ . We construct  $P_{\phi}$  likewise. We perform an element-wise addition of  $P_{\theta}$  and  $P_{\phi}$  to build the final pooled gradient matrix  $P$ :

$$P = P_{\theta} + P_{\phi} \quad (7)$$

The above pooled matrix contains gradient magnitudes of only those regions that have voxels with dominating angles in that region. Hence, it represents an approximate surface of the object with angular dominance capturing the shape information.

## 2.4 A Shape based Loss Function

Let us denote the output of 3D pooled gradient matrix from the ground truth mask as  $P_{gt}$  and that from the predicted mask as  $P_{pred}$ . Thus, the loss function can be calculated as the error/difference between  $P_{gt}$  and  $P_{pred}$ . We can represent the loss function based on the pooled gradient matrix as

$$\mathcal{L}_P = \sqrt{\|P_{gt} - P_{pred}\|^2} \quad (8)$$

The above loss function focuses on the shape of the predicted mask and penalizes on the dissimilarity of the shape of the prediction. However, we also need to focus on the overall

structure and region for correct prediction. So, we use the Dice Loss [17] along with our pooled gradient loss to build a composite function. This composite loss function is given by:

$$\mathcal{L} = \mathcal{L}_{dice} + \lambda \mathcal{L}_P \quad (9)$$

where  $\lambda$  denotes a weight with value in  $(0, 1]$  and is adjusted experimentally.

## 3 Experimental Results

In this section, we first discuss the data preparation and implementation details, followed by an ablation study and finally we show comparisons with several state-of-the-art methods. We implemented our network in PyTorch [18] and trained it on NVIDIA Titan RTX 24GB GPU. Dice Similarity Score (DSC) is used as the performance metric [17] as this is the common metric used by all other competing methods.

### 3.1 Data Preparation

Two publicly available datasets, namely, **HarP** [19] and **Kulaga-Yoskovitz** [20] were used for our experiments. Both of them contains T1-weighted MRI volumes with ground truth. The complete release of HarP includes 135 T1-weighted MRI volumes which we divided into 80% training, 10% validation, and 10% testing. **Kulaga-Yoskovitz** (KY) dataset comprises 25 participants with manually segmented labels. Each subject's MR data consists of isotropic 3D-MPRAGE T1-weighted ( $0.6mm^3$ ) and anisotropic 2D T2-weighted images. Only the former were evaluated for our investigation using a 5-fold cross validation.

Different datasets have different voxel spacings. To standardize the voxel spacing in the dataset, we registered the images and transformed the labels corresponding to MNI template ( $1 \times 1 \times 1 mm^3$  T1-w, dimensions:  $182 \times 218 \times 182$ , skull stripped) using FSL [2]. Then we skull stripped all the volumes and performed n4 bias field correction. Finally, we normalize all input images to have zero mean and unit standard deviation (std) based on non-zero voxels only. We have merged the ground truths of left and right HC into a single vol-

**Table 1:** Comparison with state-of-the-art methods

Approach	DSC		Time (in secs)
	Harp	KY	
Level Set (2019) [4]	0.847	-	-
MAS (2021) [3]	0.885	-	8.88
Hippodeep (2018) [21]	0.85	0.82	2
SWANS (2022) [22]	0.878	-	20
Hippmapp3r (2019) [5]	0.87	0.836	14
E2DHipSeg (2021) [23]	0.88	0.84	15
Subfield (2022) [24]	0.934	0.902	-
Shape SDM (2020) [9]	0.843	-	-
Shape fitting (2020) [8]	0.856	0.92	150
Our Method	<b>0.947</b>	<b>0.923</b>	12

**Table 2:** Ablation study for the proposed loss function

Loss Function	DSC	
	Harp	KY
Dice Loss	0.85	0.835
<b>Dice + Pooled Gradient Loss</b>	<b>0.947</b>	<b>0.923</b>

ume and trained our models to predict both the HC at once.

### 3.2 Implementation Details

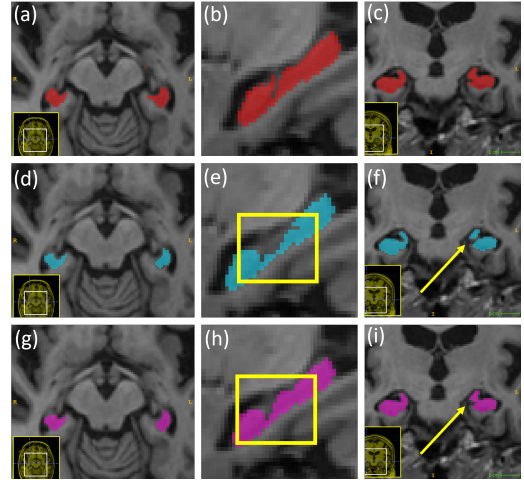
The cell size  $n$  (in Sec. 2.3) is set as 4. The parameter  $\lambda$  (Eq. 9) is gradually increased from 0.1 to 0.5 in each epoch of training. This is done to prevent the pooled gradient loss from incurring huge penalty at the beginning. We make the model learn the coarse HC region in the beginning and then the fine HC shape at a later stage. The parameter  $\beta$  (in Sec. 2.3) is set to  $\pm 12^\circ$ . For this work, we divide a circle (see Fig. 2) into 8 equal sectors and consider  $\alpha$  as the central angle. For example, if  $\alpha$  is  $22.5^\circ$ , then all angles within the range  $[22.5^\circ - 12^\circ, 22.5^\circ + 12^\circ]$  are considered. For the Attention Unet model, we used Adam optimizer with a batch size of 8. We trained the model for 200 epochs with our custom loss function (in Eq. 9).

### 3.3 Ablation Study

We have conducted an ablation study to demonstrate the effectiveness of the proposed loss function with that of the 3D Attention Unet having only Dice loss. The DSC values in Table 2 clearly indicate that our proposed loss function has achieved superior results on both the datasets over the Dice loss alone. Qualitative results for a sample from the HarP dataset, as shown in Fig. 3, corroborate the same.

### 3.4 Comparisons with Other Methods

We have compared our method with as many as nine state-of-the-art segmentation approaches from three different categories and the results are shown in Table 1. Among these nine methods, two did not use any DL (marked in blue), five used DL with regular loss functions like Dice or cross-entropy (marked in green) and two applied DL with shape based loss



**Fig. 3:** Visual comparisons of ground truth segmentation with a model having only dice loss and the same model having the proposed loss function: (a), (b) and (c) represent the axial, sagittal and coronal views of the ground truth HC segmentation, (d), (e) and (f) represent the same views of the predicted segmentation output of the model with dice loss only, (dice score: 0.848), (g), (h) and (i) represent the same views of the predicted segmentation of our proposed loss function (dice score: 0.927). The difference in prediction can be seen in (e) vs. (h) and in (f) vs. (i) and is highlighted with yellow squares and yellow arrows respectively.

functions (marked in red). Some methods like Hippodeep, Hippmapp3r, E2DHipSeg and Subfield have used multiple datasets. Hence, we re-implemented those works using HarP and KY datasets with their publicly available codes. However, we left some blank entries in the KY column as the corresponding works did not report Dice Scores for the KY dataset and had no publicly available code for re-implementation. Results in the Table 1 clearly establish that the proposed method achieves state-of-the-art results by outperforming all other competing methods and ranking third best in terms of the execution time.

## 4 Conclusion

Automated segmentation of the HC is found to be immensely helpful in improving throughput of neuro-radiologists as well as in fast treatment planning for neuro-surgeons. We presented a HOG based loss function in Attention UNet for 3D segmentation of HC, a complex irregular structure in brain. We also designed a pooling function to address the derivability issue of histogram in HOG. In the future, we plan to incorporate shape based loss function to accurately detect structures of varying shapes and sizes. Another direction of future research will be to further increase performance by using state-of-the-art backbone segmentation models.

## Acknowledgement

Arijit De was supported by Tata Consultancy Services (TCS) Research Scholar Program (RSP) of TCS Pvt. Ltd.

## 5 References

- [1] C. La et al., “Hippocampal cal subfield predicts episodic memory impairment in parkinson’s disease,” *NeuroImage: Clinical*, vol. 23, pp. 101824, 2019.
- [2] B. Fischl, “Freesurfer,” *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.
- [3] W. Wang et al., “A robust discriminative multi-atlas label fusion method for hippocampus segmentation from mr image,” *Computer Methods and Programs in Biomedicine*, vol. 208, pp. 106197, 2021.
- [4] N. Safavian, S. A. H. Batouli, and M. A. Oghabian, “An automatic level set method for hippocampus segmentation in mr images,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 8, no. 4, pp. 400–410, 2020.
- [5] M. Goubran et al., “Hippocampal segmentation for brains with extensive atrophy using three-dimensional convolutional neural networks,” Tech. Rep., Wiley Online Library, 2020.
- [6] S. Suganyadevi, V. Seethalakshmi, and K. Balasamy, “A review on deep learning in medical image analysis,” *International Journal of Multimedia Information Retrieval*, vol. 11, no. 1, pp. 19–38, 2022.
- [7] Q. Huang, Y. Zhou, and L. Tao, “Dual-term loss function for shape-aware medical image segmentation,” in *ISBI*. IEEE, 2021, pp. 1798–1802.
- [8] I. Brusini et al., “Shape information improves the cross-cohort performance of deep learning-based segmentation of the hippocampus,” *Frontiers in Neuroscience*, vol. 14, pp. 15, 2020.
- [9] H. Tang et al., “Shape-aware organ segmentation by predicting signed distance maps,” 2022, US Patent 11,301,999.
- [10] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *CVPR*. IEEE, 2005, vol. 1, pp. 886–893.
- [11] B. Bhattarai et al., “Histogram of oriented gradients meet deep learning: A novel multi-task deep network for 2d surgical image semantic segmentation,” *Medical Image Analysis*, vol. 85, pp. 102747, 2023.
- [12] A. Serag et al., “Histograms of oriented 3d gradients for fully automated fetal brain localization and robust motion correction in 3 t magnetic resonance images,” *BioMed Research International*, vol. 2017, 2017.
- [13] Ö. Çiçek et al., “3d u-net: learning dense volumetric segmentation from sparse annotation,” in *MICCAI*. Springer, 2016, pp. 424–432.
- [14] O. Oktay et al., “Attention u-net: Learning where to look for the pancreas,” in *Medical Imaging with Deep Learning*, 2018.
- [15] J. Kugelman et al., “A comparison of deep learning u-net architectures for posterior segment oct retinal layer segmentation,” *Scientific Reports*, vol. 12, no. 1, pp. 14888, 2022.
- [16] P. Scovanner, S. Ali, and M. Shah, “A 3-dimensional sift descriptor and its application to action recognition,” in *Proceedings of the 15th ACM international conference on Multimedia*, 2007, pp. 357–360.
- [17] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [18] A. Paszke et al., “Pytorch: An imperative style, high-performance deep learning library,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [19] M. Boccardi et al., “Training labels for hippocampal segmentation based on the eadc-adni harmonized hippocampal protocol,” *Alzheimer’s & Dementia*, vol. 11, no. 2, pp. 175–183, 2015.
- [20] K. Yoskovitz et al., “Multi-contrast submillimetric 3 tesla hippocampal subfield segmentation protocol and dataset,” *Scientific data*, vol. 2, no. 1, pp. 1–9, 2015.
- [21] B. Thyreau, K. Sato, H. Fukuda, and Y. Taki, “Segmentation of the hippocampus by transferring algorithmic knowledge for large cohort processing,” *Medical Image Analysis*, vol. 43, pp. 214–228, 2018.
- [22] E. Balboni et al., “The impact of transfer learning on 3d deep learning convolutional neural network segmentation of the hippocampus in mild cognitive impairment and alzheimer disease subjects,” *Human Brain Mapping*, vol. 43, no. 11, pp. 3427–3438, 2022.
- [23] D. Carmo et al., “Hippocampus segmentation on epilepsy and alzheimer’s disease studies with multiple convolutional neural networks,” *Heliyon*, vol. 7, no. 2, pp. e06226, 2021.
- [24] J. V. Manjón, J. E. Romero, and P. Coupe, “A novel deep learning based hippocampus subfield segmentation method,” *Scientific Reports*, vol. 12, no. 1, pp. 1333, 2022.