Hippocampus Segmentation Based on a Modified UNet Architectures

Baraa Rayed[1] and Koray Çiftçi[2]

[1,2] *Biomedical Engineering Department, Tekirdağ Namık Kemal University*

February 14, 2023

Abstract

Background and aim: The investigation, diagnosis, and choice of treatment for neuropsychiatric illnesses depend critically on the segmentation of the hippocampus from magnetic resonance imaging data. Deep learning is now widely used in various models for automatic segmentation. Our aim is to develop a deep learning-based methodology for hippocampus segmentation.

Method: In this research, we developed an open-source hippocampus segmentation method based on deep learning. The method employs automatic preprocessing and utilizes two 2D neural networks. A publicly available data-set for the segmentation of the hippocampus in Alzheimer's disease was used to develop and test the methodology.

Results: The effectiveness of our two models for 2D segmentation of the hippocampus was demonstrated by the training and testing performances of the algorithms on the data-set. The modified models (U-Net and DC-U-Net) achieved high accuracy scores of 0.98 and 0.99, respectively, on testing using 5-fold cross-validation.

Keywords: Alzheimer's disease, Deep Learning, Hippocampus, MRI, Segmentation.

1 Introduction

Investigation of hippocampal formation is a first step in many neurological disorder's like Alzheimer's Disease, and epilepsy [1]. Correct segmentation of the hippocampus is thus crucial for an accurate diagnosis and treatment planning [2]. Manual segmentation of the hippocampus is time-consuming and requires specialized knowledge. The need for efficient automatic segmentation techniques has been sparked by the high workload of manual segmentation [3]. A manual finer segmentation is already employed as a starting point for some of those techniques, such as FreeSurfer[4].

The majority of the most recent hippocampus segmentation techniques involve convolutional neural networks (CNN) [5]. However, a caveat regarding these studies is that many of these algorithms can only access healthy scans or patients with Alzheimer's disease because they rely on publicly accessible datasets for training and evaluation. This raises the worry that automated techniques might only be equipped to handle features found in publicly available datasets on Alzheimer's disease and healthy participants, including ADNI and the Multi Atlas Labeling Challenge (MALC)[5]. In the current study, we present a deep learning based hippocampus segmentation method equipped with tuning and data augmentation. Our method employs data augmentation techniques, such as rotation, scaling, and flipping to artificially increase the size of the dataset and improve the robustness of the model[6]. We also use domain adaptation techniques to fine-tune our model on a smaller dataset of the target domain. This helps to improve the generalization capabilities of our algorithm and reduce the overfitting of the model to the training data.

1.1 Contributions

We offer an easily accessible approach for segmenting the hippocampus using a group of 2D CNNs. The method achieves a high level of accuracy on HarP public data and makes use of current developments in the deep learning literature.

This paper is organized as follows: Section 2 provides an overview of the neural networks used in the study. Section 3 presents a literature review of current deep learning-based hippocampus segmentation techniques. Section 4 details the HarP dataset utilized in the study. Our proposed hippocampus segmentation approach is explained in Section 5. The experimental results and comparisons to other methods on the HarP dataset are presented in Section 6. Discussion and further analysis of the results can be found in Section 7. Finally, the paper concludes in Section 8.

2 Understanding Neural Networks

U-Net is a type of CNN architecture that is commonly used for image segmentation tasks[7]. It consists of two main parts: a contracting path and an expansive path. The contracting path is composed of multiple convolutional and max pooling layers that reduce the spatial dimension of the input image while increasing the number of feature maps. The expan-

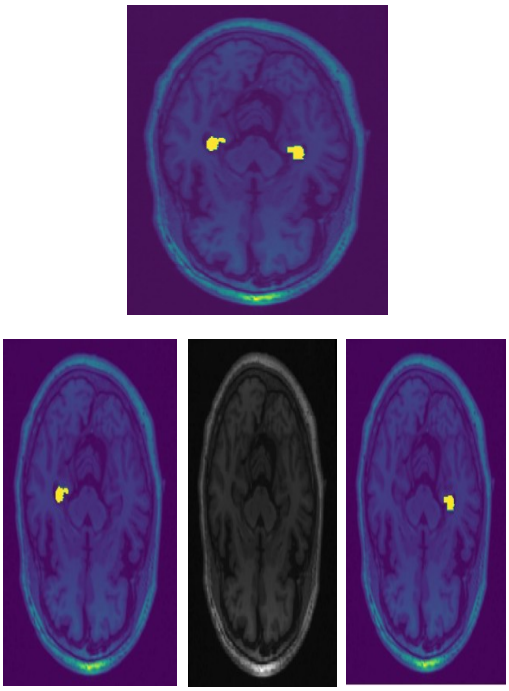

Figure 1: A manual annotation in yellow is shown for one image from the HarP dataset. The other three images show one slice from the HarP dataset with the left and right annotations in yellow.

sive path, on the other hand, is composed of multiple transposed convolutional layers that increase the spatial dimension of the feature maps while reducing the number of feature maps. The output of the contracting path is concatenated with the output of the corresponding layer in the expansive path, resulting in a more detailed segmentation of the input image. The final output of the U-Net is a probability map, where each pixel in the output represents the probability of that pixel belonging to a specific class or object in the input image.

A Deeply-supervised Conditional UNet (DC-UNet) is an extension of the U-Net architecture for medical image segmentation. It features the addition of a deeply-supervised mechanism which aims to improve the accuracy and robustness of the model.

The DC-UNet architecture is composed of a number of U-Net like layers, with the addition of auxiliary output layers placed at various depths of the network. These auxiliary output layers are connected to the corresponding layers in the contracting path and are trained to predict the segmentation maps. These auxiliary outputs provide additional supervision signals to the network, allowing the model to better learn the features of the input image.

The DC-UNet also use a conditional mechanism, where the model is conditioned on the input image, and the segmentation maps are conditioned on both the input image and the current segmentation maps. The conditioning on the input image and the segmentation maps helps the network to better capture the context of the image and make more accurate predictions.

In summary, the DC-UNet is an enhanced version of the U-Net architecture that uses deeply-supervised and conditional mechanisms to improve the accuracy and robustness of the model for medical image segmentation tasks[8].

3 Hippocampus segmentation with deep learning

Recent research on the subject of hippocampus segmentation using deep learning is examining various task-related architectures, loss functions, and general methodologies. One strategy that is used in the majority of these studies is to combine patches as training inputs with 2D or 3D CNNs. It should be noted that some efforts focus on hippocampus segmentation, whereas others aim to segment various neuroanatomies. Here is a quick synopsis of each of those works.

A group of researchers named Chen et al. [9] report 0.9 Dice [10] in 10-fold 110 ADNI [2] take a way out to talk about using the Convolutional Neural Networks CNN to measure the size of a part of the brain called the hippocampus[7]. They used a big set of pictures of the brain from a group of people with a disease called Alzheimer's. Instead of only looking at 3 views of the brain, they looked at 9 views by cutting the brain in different ways. They used 9 small computer programs called U-Nets to look at each view. They put all the results together to get the final answer.

A voxel-wise classification approach is adopted by Xie et al. [11] utilizing triplanar patches that cross the target voxel. In addition to the usual application of ReLU activations and softmax, they combine features from all patches into a deep neural network with a fully connected classifier [12]. The training patches exclusively come from the roughly center region of the hippocampus, which balances labels for target voxels in the foreground and background at a 1:1 ratio. While fully convolutional neural networks are still faster than multi-atlas methods, voxel classification techniques are typically slower. The U-Net architecture has the potential to segment data using a collection of 2D U-Nets rather than a single 3D one. For comparison in this paper's dataset, which includes a 3D UNet architecture, several of those methods were replicated [5].

4 Data

The ADNI Alzheimer's study's HarP public dataset, widely used as a benchmark for hippocampus segmentation, served as the primary source of data for this study. Our study offered an automatic method for converting the MNC files of this dataset into a series of PNG files for each patient The method preserves the ID of each file, allowing for easy recognition of the corresponding mask for each slice, as shown in (Fig. 1).

There are 135 T1-weighted MRI volumes in the complete HarP release. Control normal (CN), mild

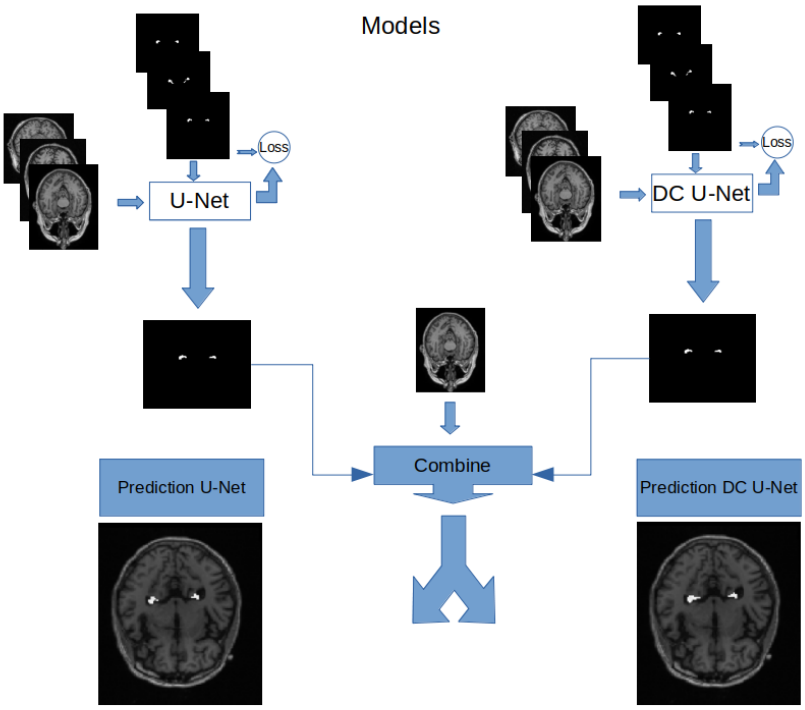

Figure 2: The final segmentation is generated by combining the original image with the prediction generated by FCNNs on 2D orientation for each slice.

cognitive impairment (MCI), and Alzheimer's disease (AD) cases all occur at almost comparable rates in the different groups of Alzheimer's disease [2]. No volumes were eliminated, and volumes were adjusted for min-max intensity between 0 and 1. While k-Folds, when used, consisted of 5 folds with no overlap on the test sets, training with stratified hold-out was carried out with 60% training, 20% validation, and 20% testing.

5 Segmentation methodology

The general methodology for our hippocampus segmentation method[1] is described in this section (see Fig. 2). We use 2D U-Net CNN models on 2D axial scans. A combination process combines the original image with the prediction generated by the CNNs. Slice by slice, the activations of each network are constructed for a specific input volume.

The design of our network was inspired by the U-Net fully convolutional neural network (FCNN) architecture. However, we made several adjustments to the architecture based on previous successful works [13] (Fig. 3). These changes included the addition of batch normalization layers in the convolution block for the U-Net architecture and the modification of the first layers of the DC block with an alpha value for the DC-U-Net. Each DC block consists of six convolutional layers with two concatenation layers and an add layer (Fig. 4). The ResPath refers to the path through the network used to calculate the residuals of the DC-U-Net, which involves calculating the residuals at each layer of the network and using them to update the weights of the network in an attempt to improve its performance. It's important to note that the term ResPath is not a standard term in the field of deep learning, and its specific meaning and implementation may vary depending on the context in which it is used. These two blocks were used to construct the DC-U-Net architecture based on the standard U-Net architecture(Fig. 5)[14].

Slices for each network are taken throughout prediction time using a center crop. The resultant activation is padded back to the original size when taking the predictions produced by the networks and combining them back with the original slices. The slices and their masks are inputted orderly to the two models with the same size for each of them in the training. For testing, different sets are used to test the models with the same shape as the training set.

5.1 Loss Function

In the context of a segmentation task, binary cross-entropy loss is a measure of how well the model is able to predict the correct class for each pixel in an image. It is commonly used when there are only two classes (e.g., object and background). The binary cross-entropy loss function is defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (1)$$

where N is the total number of pixels in the im-

[1]Source code: https://github.com/Baraa-Rayed/Hipseg-mod-Unets

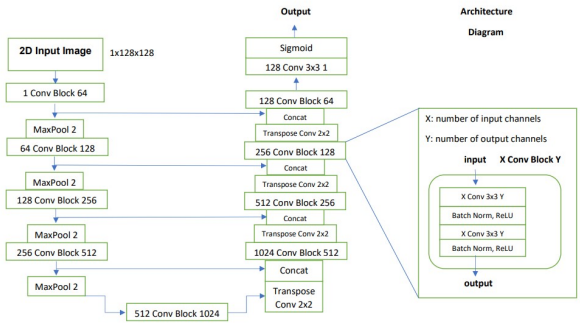

Figure 3: the Final architecture of modified U-Net in Fig. 2. in comparison with original U-Net [15] use of BatchNorm in the Conv Block, with the sigmoid Fully connected layers output limitation

.

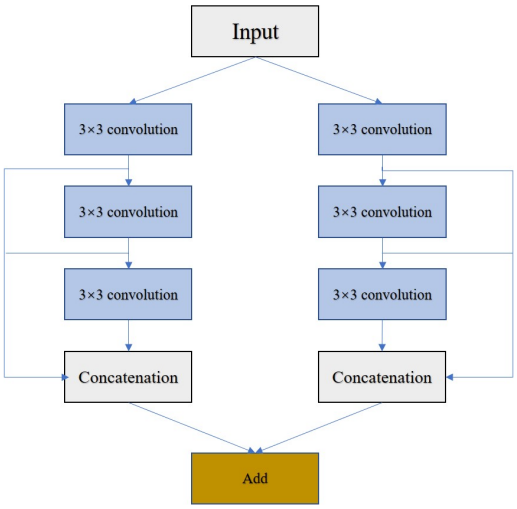

Figure 4: DC Block: Essential Element of a DC U-Net for Image Segmentation[14].

age, y_i is the true label (0 or 1) for pixel i, and $\hat{y}_i$ is the predicted probability that pixel i belongs to the object class (also 0 or 1). The loss function penalizes the model for predicting the wrong class, with a larger penalty for confident incorrect predictions (i.e., when $\hat{y}_i$ is close to 1 for a negative example, or $\hat{y}_i$ is close to 0 for a positive example). During training, the model adjusts its parameters to minimize the binary cross-entropy loss, which encourages the model to predict probabilities that are close to the true labels and leads to more accurate segmentation[16].

The loss function is calculated as the negative log-likelihood of the true label, given the predicted probability. In other words, it measures how well the model's predicted probability aligns with the true label.

The binary cross-entropy loss is often used in conjunction with the sigmoid activation function, which maps the output of a model to a value between 0 and 1. This allows the model to predict the probability of the label being 1.6

In summary, the loss function used in a segmentation task is a measure of how well a model's predictions align with the ground truth masks. It is used to update the model's weights and biases during training in order to improve its performance. In the case of segmentation, the loss function is typically calculated based on the pixels in the predicted mask and the corresponding pixels in the ground truth mask.

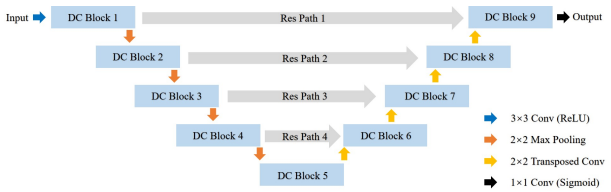

Figure 5: DC unet, which is a convolutional neural network (CNN) designed for image segmentation tasks[14].

5.2 Data augmentation and hyperparameter tuning

We improved the performance of the two models by tuning the hyperparameters manually. To enhance the data, we used the image data generator technique which involved flipping the images and masks horizontally, applying a 0.1 range for zooming, shift widths and shift heights, and a 15 range for rotation. These augmentations were applied to all three sets (training, validation, and test) to ensure the models would perform well and to prevent overfitting or underfitting problems.

5.3 Combine

Our approach involves combining the left and right labels before inputting the data into two models from the U-Net family. The prediction masks produced by each model are then combined with the original image. By comparing the results of these two models, we aim to achieve the goal of this paper as illustrated in (Fig. 2).

6 Results

In this section, we present a comparative analysis of the performance of two segmentation models on the HarP dataset. The results are presented quantitatively in a table, which shows the performance of the models with different variations on the HarP test set.

6.1 Qualitative Outcomes

Visual inspection of the HarP data revealed very little variance. We didn't find any heavy outliers, which means that there were no extreme values that are significantly different from the rest of the values in the dataset. Even though the HarP dataset is still poorly giving all the information to distinguish between the diseases inside each slide, it is still useful for our goal

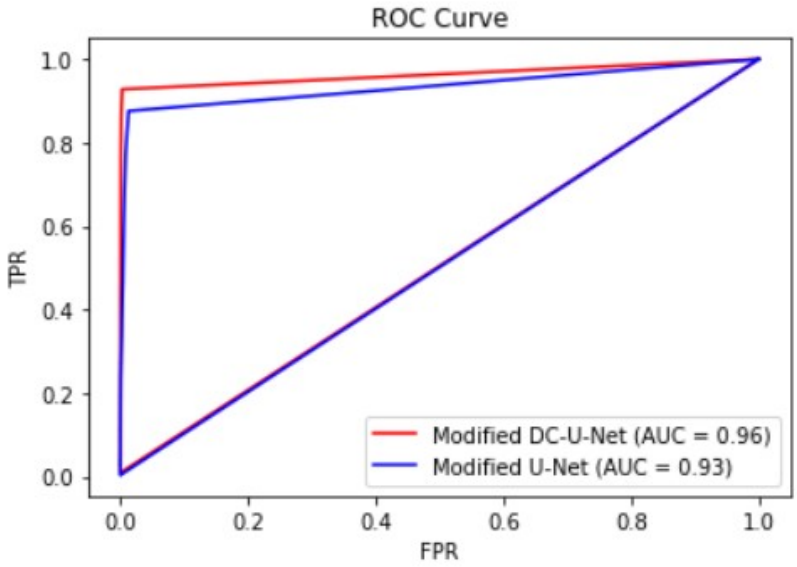

Figure 6: Evaluating Segmentation Performance using Receiver Operating Characteristic Curves for Modified U-Net and Modified DC-U-Net.

of attempting to modify and use the modern neural network DC-U-Net and the classic neural network U-Net from the U-Net family.

During the evaluation phase, we initially trained the two models without making any changes to their architectures. However, the models did not achieve high accuracy. Therefore, we faced the challenge of making modifications to the models' structures or using different hyperparameters to compile the models in order to improve their performance. After several attempts to improve the performance of the models, we were able to generate two versions that performed exceptionally well. These versions included the addition of two batch normalization layers to the Conv Block of the classic U-Net, as well as the modification of the first layer of the DC-Block with different tunes to the hyperparameters.

6.2 Quantitatively Results

To compare the performances of these two models, we used data augmentation and hyperparameters tuning to enhance the performance of the models on the same dataset, AS we divided our dataset into three sets: training, validation, and testing. To enhance the dataset, we applied augmentation techniques, such as rotation, cropping, flipping, and zooming, on some examples from these sets. Additionally, we fine-tuned the hyperparameters, including the learning rate, batch size, and the number of epochs. As a result of these modifications and techniques, We present the ROC curve for the Modified U-Net and Modified DC-U-Net in (Fig. 6).

Both models showed outstanding results in segmenting the hippocampus by adjusting the threshold between 0 and 1 to evaluate TPR (True Positive Rate) and FPR (False Positive Rate) between the predictions and the annotations. TPR, also known as Sensitivity, is the ratio of correctly classified positive cases to all positive cases. FPR, or fall-out, is the ratio of incorrectly classified positive cases to all negative cases. This is represented in the ROC (Receiver Operating Characteristic) curve[17], with a high TPR indicating a successful model. It is worth noting that these two models outperformed the other available models for this task(Table 1).

7 Discussion

The Consensus technique effectively removes the majority of false positives that are typically generated by some networks. This technique involves averaging the activations from the networks, followed by thresholding and post-processing. After the construction of the consensus of activations, a threshold is needed to binarize the segmentation. Small structures of the brain similar to the hippocampus could be classified as false positives. To remove those false positives, a 3D labeling implementation from [18] was used, with subsequent removal of small non-connected volumes, keeping the 2 largest volumes, or 1 if a second volume is not present. This post-processing step is performed after the average consensus of all networks and threshold applications.

As future work, this strategy can be further applied by using the DC-U-Net from our method for 2D multi-orientation hippocampus segmentation on a new dataset that includes axial, sagittal, and coronal planes with their annotations. By averaging the activations from the three CNNs, the Consensus strategy enhances the accuracy of hippocampus segmentation without being affected by minor false positives in other parts of the brain. Additionally, it was observed that when one of the networks fails, the consensus of the other two networks can "save" the outcome.

Utilizing the DC-U-Net as multiple networks on

Segmentation Models	
Deep Learning Models	HarP Dataset
Modified U-Net	0.98 (this work)
Modified DC-U-Net	0.99 (this work)
E2DHipseg(2021)[5]	0.90
Ataloglou et al.(2019)[13]	0.90

Table 1: Reported testing results for HarP.This work in the literature has been named as Hipseg-mod-Unets. Results were calculated following 5-fold cross-validation

the three planes axial, coronal, and sagittal with a consensus strategy resulted in improved segmentation of the hippocampus. This is because the DC-U-Net demonstrated higher accuracy than the U-Net in 2D segmentation. It's important to note that the sample size on each plane can affect the performance of the model, and thus, the sample sizes should be similar across all planes. Additionally, this study would benefit from implementing a data preprocessing method that saves the identification of each image for the 3D labeling.

The process of segmenting the hippocampus could potentially be simplified in the future through the use of a 3D DC-U-Net instead of multiple 2D U-Nets with a consensus strategy. The focus would be on improving a single model, which could be achieved by adjusting the alpha values of the DC-block or adding additional layers. This approach could be applied to data with multiple classes, such as the HUCnicamp data, and evaluated using a logical loss function such as dice loss. The use of dice loss aims to achieve the goals of the study in a more efficient and innovative manner.

8 Conclusions

This study aimed to improve the performance of two segmentation models from the U-Net family, the DC-U-Net and the U-Net, on the hippocampus segmentation task using the HarP dataset. We used various optimization techniques for this purpose. Our findings demonstrated that the DC-U-Net, which combines the U-Net architecture with the DC Block, outperformed the U-Net in terms of both efficiency and accuracy. These results suggest that the DC-U-Net is a more effective model for segmentation tasks compared to the original U-Net.

Our segmentation models have proven highly effective in this task, as they are able to accurately identify even the small pieces of the hippocampus in the first and last slices that contain it.

There are concerns that existing automatic hippocampus segmentation techniques may not be suitable for handling cases where the hippocampus has been removed due to epilepsy treatment. However, our research has demonstrated the benefits of using modern segmentation techniques to label the hippocampus, as this can help researchers to save time and do more efficiently study to find new treatments for neurological disorders such as lobe epilepsy.

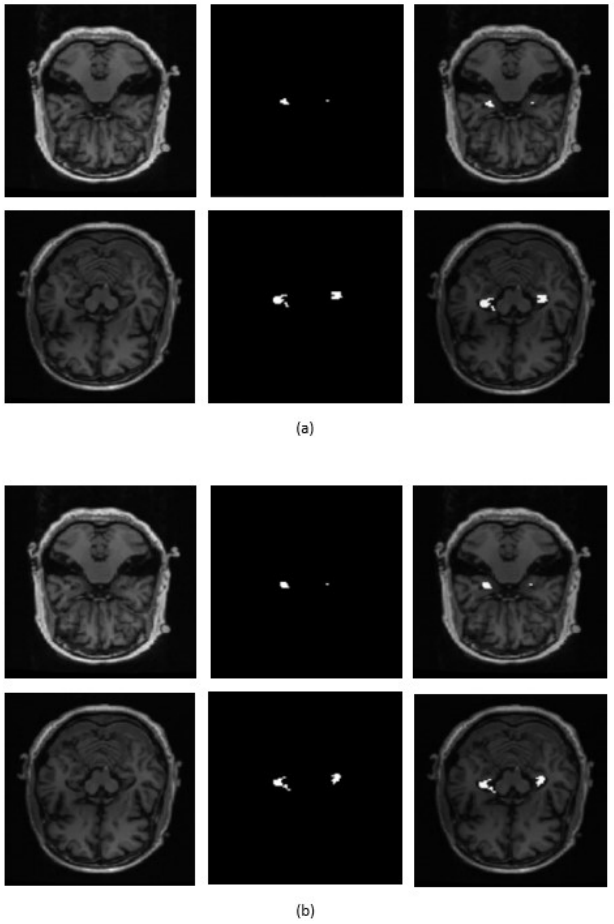

Figure 7: Multiview for HarP dataset images with their labels (a) and our models prediction on it (b) .

References

[1] P. Andersen, R. Morris, D. Amaral, T. Bliss, and J. O'Keefe, *The hippocampus book* (Oxford university press, 2006).

[2] E. Ghizoni, J. Almeida, A. F. Joaquim, C. L. Yasuda, B. M. de Campos, H. Tedeschi, and F. Cendes, Modified anterior temporal lobectomy: anatomical landmarks and operative technique, *Journal of Neurological Surgery Part A: Central European Neurosurgery* **76**, 407 (2015).

[3] C. McCarthy and T. Vaughan, in *Numerical Modelling of Failure in Advanced Composite Materials* (Elsevier, 2015), 379–409.

[4] B. Fischl, Freesurfer, *Neuroimage* **62**, 774 (2012).

[5] D. Carmo, B. Silva, C. Yasuda, L. Rittner, R. Lotufo, A. D. N. Initiative, et al., Hippocampus segmentation on epilepsy and alzheimer's disease studies with multiple convolutional neural networks, *Heliyon* **7**, e06226 (2021).

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Communications of the ACM* **60**, 84 (2017).

[7] O. Ronneberger, P. Fischer, and T. Brox, in *International Conference on Medical image computing and computer-assisted intervention*

(Springer, 2015), 234–241.

[8] L. Huang, T. Song, K. Yu, F. Yuan, H. Wang, J. Zhi, G. Hu, and H. Yang, in *International Conference on Artificial Intelligence and Intelligent Information Processing (AIIIP 2022)* (SPIE, 2022), vol. 12456, 548–553.

[9] Y. Chen, B. Shi, Z. Wang, P. Zhang, C. D. Smith, and J. Liu, in *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)* (IEEE, 2017), 192–196.

[10] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, in *Deep learning in medical image analysis and multimodal learning for clinical decision support* (Springer, 2017), 240–248.

[11] Z. Xie and D. Gillies, Near real-time hippocampus segmentation using patch-based canonical neural network, *arXiv preprint arXiv:1807.05482* (2018).

[12] A. Krizhevsky, I. Sutskever, and G. Hinton, Imagenet classification with deep convolutional neural networks. 2012 advances in neural information processing systems (nips), *Neural Information Processing Systems Foundation, La Jolla, CA* (2012).

[13] D. Ataloglou, A. Dimou, D. Zarpalas, and P. Daras, Fast and precise hippocampus segmentation through deep convolutional neural network ensembles and transfer learning, *Neuroinformatics* **17**, 563 (2019).

[14] A. Lou, S. Guan, and M. Loew, in *Medical Imaging 2021: Image Processing* (SPIE, 2021), vol. 11596, 758–768.

[15] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2020), 1055–1059.

[16] F. van Beers, A. Lindström, E. Okafor, and M. A. Wiering, in *ICPRAM* (2019), 438–445.

[17] J. A. Hanley and B. J. McNeil, The meaning and use of the area under a receiver operating characteristic (roc) curve., *Radiology* **143**, 29 (1982).

[18] E. R. Dougherty and R. A. Lotufo, *Hands-on morphological image processing*, vol. 59 (SPIE press, 2003).