Hippocampus Substructure Segmentation Using Morphological Vision Transformer Learning

Yang Lei[1], Yifu Ding[1], Richard L.J. Qiu[1], Tonghe Wang[2], Justin Roper[1], Yabo Fu[2], Hui-Kuo Shu[1], Hui Mao[3] and Xiaofeng Yang[1*]

1Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA 30308

2Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, 10065

3Department of Radiology and Imaging Sciences and Winship Cancer Institute, Atlanta, GA 30308

*Corresponding author:

Xiaofeng Yang, PhD

Department of Radiation Oncology

Emory University School of Medicine

1365 Clifton Road NE

Atlanta, GA 30322

E-mail: xiaofeng.yang@emory.edu

Abstract

Background: The hippocampus plays a crucial role in memory and cognition. Because of the associated toxicity from whole brain radiotherapy, more advanced treatment planning techniques prioritize hippocampal avoidance, which depends on an accurate segmentation of the small and complexly shaped hippocampus. Purpose: To achieve accurate segmentation of the anterior and posterior regions of the hippocampus from T1 weighted (T1w) MRI images, we developed a novel model, Hippo-Net, which uses a mutually enhanced strategy.

Methods: The proposed model consists of two major parts: 1) a localization model is used to detect the volume-of-interest (VOI) of hippocampus. 2) An end-to-end morphological vision transformer network is used to perform substructures segmentation within the hippocampus VOI. The substructures include the anterior and posterior regions of the hippocampus, which are defined as the hippocampus proper and parts of the subiculum. The vision transformer incorporates the dominant features extracted from MRI images, which are further improved by learning-based morphological operators. The integration of these morphological operators into the vision transformer increases the accuracy and ability to separate hippocampus structure into its two distinct substructures.

A total of 260 T1w MRI datasets from Medical Segmentation Decathlon dataset were used in this study. We conducted a five-fold cross-validation on the first 200 T1w MR images and then performed a hold-out test on the remaining 60 T1w MR images with the model trained on the first 200 images. The segmentations were evaluated with two indicators, 1) multiple metrics including the Dice similarity coefficient (DSC), 95th percentile Hausdorff distance (HD95), mean surface distance (MSD), volume difference (VD) and center-of-mass distance (COMD); 2) Volumetric Pearson correlation analysis.

Results: In five-fold cross-validation, the DSCs were 0.900±0.029 and 0.886±0.031for the hippocampus proper and parts of the subiculum, respectively. The MSD were 0.426±0.115mm and 0.401±0.100 mm for the hippocampus proper and parts of the subiculum, respectively. Conclusions: The proposed method showed great promise in automatically delineating hippocampus substructures on T1w MRI images. It may facilitate the current clinical workflow and reduce the physicians' effort.

Keywords: hippocampus substructure, segmentation, deep learning

1 INTRODUCTION

The hippocampus is a pair of medial and subcortical brain structures located in proximity to the temporal horn of the lateral ventricles, which is an active research area due to its implication in memory and neuropsychiatric disorders.[14] In radiation therapy, hippocampal avoidance whole brain radiation using volumetric modulated arc therapy (VMAT) plus the medication memantine has been shown to preserve cognitive function without compromising progression-free survival or overall survival when compared to classic whole brain radiation therapy plus memantine.[3, 2] In Alzheimer's Disease (AD), the progression of AD occurs from the trans-entorhinal cortex to the hippocampus, and finally to the neocortex.[2] These progression steps depend on the severity of the neurofibrillary tangles found in neuropathological studies. However, similar patterns can also be observed in the progress of brain atrophy found on MRI imaging studies. The atrophy of hippocampus measured from MRIs can be used as an early sign of AD progression.[1] Additionally, evidence of hippocampal atrophy as measured from MRIs can occur before the onset of clinical symptoms.[14] Therefore, accurate segmentation of the hippocampus from MRIs is a meaningful task in medical image analysis across multiple disciplines.[26]

To determine whether the hippocampus is atrophic, clinicians often need to segment the bilateral hippocampus on MRI scans and analyze their shape and volume.[19, 27] This task is difficult, however, due to several factors. Firstly, the hippocampus has low contrast with the surrounding tissues on MRI scans,[25] since it is a gray matter structure. Secondly, the hippocampus has an irregular shape leading to a blurred boundary in cross-sectional slices.[5] Thirdly, the hippocampus is a small structure with limited volume as compared to other structures that are routinely delineated as organs-at-risk (OARs) in radiation therapy.[8] Finally, there are large variations in the size and shape of the hippocampus across patients.[4] Therefore, accurate and automatic segmentation of hippocampus is a challenging task. Until now, manual segmentation of hippocampus is still the standard in clinical practice.[13] However, manual segmentation is a tedious and error-prone process, which limits its application in big data and clinical practice. Thus, many efforts have been devoted to developing computer-aided diagnostic systems for automated segmentation of the hippocampus.

The existing automatic hippocampal segmentation methods can be categorized into two main types: atlas-based methods and machine learning-based methods. Atlas-based methods can be further divided based on the number of atlases used in the segmentation process into single-atlas-based, average-shape atlas-based, and multi-atlas-based approaches. For instance, Haller et al. first proposed to use the single-atlas-based approach for hippocampal segmentation.[11, 10] However, single-atlas-based approaches are limited by inter-patient variations. To address this, average shape-based mapping approaches were proposed to overcome such limitations, but the segmentation results depend on the alignment quality of the target and average maps. Thus, a priori knowledge of medical mapping was incorporated into the multi-atlas-based segmentation approach. For example, Wang et al. proposed a robust discriminative multi-atlas label fusion approach to segment hippocampus by building the conditional random field (CRF) model that combines distance metric learning and graph cuts.[28] Wang's approach is a patch embedding multi-atlas label fusion method that utilizes only the relationship between the target block and the atlas block, and ignores the possibility that unrelated atlas blocks may dominate the voting process. Existing atlas-based methods do not consider the anatomical differences in hippocampus among patients, and do not consider the correlation between atlases.

Machine learning-based methods can be further classified into traditional machine learning-based approaches and deep learning-based approaches. Traditional machine learning-based approaches mainly include support vector machine (SVM), Markov random field (MRF), principal component analysis (PCA), et al.[15, 17] For instance, Hao et al. proposed a local label learning strategy to estimate segmentation labels of target images by using SVM with image intensity and texture features.[12] However, these traditional approaches to machine learning rely heavily on the quality of handcrafted features, and further suffer from slow segmentation, susceptibility to noise interference, and insufficient generalization performance.[16]

Because convolutional neural network (CNN) models can automatically extract the pixel feature information from images, they have been widely used in multiple medical image analysis tasks.[9] For example, CNN-based models can be used to segment the hippocampus from MRIs.[20] Qiu et al. proposed a multitask 3D U-net framework for hippocampus segmentation by minimizing the difference between the targeted binary mask and the model prediction, and optimizing an auxiliary edge-prediction task.[23] Cao et al. developed a two-stage segmentation method to perform the task of 3D hippocampus

segmentation by localizing multi-size candidate regions and fusing the multi-size candidate regions.[6] These methods show promising results, demonstrating the potential of CNN-based models to improve the efficiency and accuracy of hippocampus segmentation.

However, most existing deep learning-based methods ignore the spatial information of the hippocampus relative to the entirety of the human brain. As a result, they cannot effectively fuse the shape features and the semantic features, which leads to lower segmentation accuracy. Hippocampal tracing began from anterior where the head is visible as an enclosed gray matter structure inferior to the amygdala, and continued posteriorly using surrounding white matter or CSF as boundaries. Subiculum (posterior parts of hippocampus) was included in the hippocampus. Delineation stopped when the wall of the ventricle was visibly contiguous with the fimbria. The subiculum occupies a portion of the para-hippocampal gyrus in the mesial temporal lobe and is a component of the medial temporal memory system. Therefore, in this work, we aim to develop a novel deep network framework to segment the hippocampus by introducing a spatial attention mechanism to capture the spatial location information of the hippocampus relative to the brain. We also designed a cross-layer dual encoding shared decoding network to extract the semantic characteristics of the hippocampus. By combining the spatial location information and semantic characteristics of the hippocampus, we enhanced the segmentation accuracy of the hippocampus. In this study, we trained a novel morphological visual transformer learning-based hippocampus substructure segmentation for accurate segmentation of the anterior and posterior regions of the hippocampus from T1 weighted (T1w) MR images.

2 Methods and Materials

2.1 Overview

Figure 1 outlines the schematic flow chart of this hippocampus multi-substructure segmentation process. The proposed network follows the same feedforward path for both training and inference. A collection of hippocampus images and multi-substructure contours was used for model training. The proposed model, named as morphological visual transformer-based network, takes the hippocampus image as input and generates the auto-contour of two substructures, which are the hippocampus proper and parts of the subiculum. The manual contours of these two substructures were used as ground truth to supervise the proposed network.

The proposed model consists of two deep learning-based subnetworks, i.e., a localization model and a segmentation model. The localization model is a hippocampus) detection network that is used to detect the volume-of-interest (VOI) for both the hippocampus proper and parts of the subiculum[7] from the T1w MR image. The MR image is then cropped within the VOI before transfer to the segmentation subnetwork to ease the computational task. The segmentation model is implemented via an end-to-end morphological vision transformer network, which is used to perform substructures segmentation within the hippocampus VOI. The vision transformer incorporates the dominant features extracted from MR images. The integration of the morphological operators into the vision transformer increases the ability of separating the hippocampus into two substructures.

During inference, the trained localization model takes a hippocampus T1w MR image as input and detects the VOI of hippocampus as the first step. Then, the cropped image within the VOI is sent to the segmentation model, i.e., morphological visual transformer, to segment the substructures. Finally, based on the detected coordinates derived by the localization model, the segmented contour is converted back to its original coordinates to obtain the final segmentation.

2.2 Localization model

The aim of the localization model is to crop the image to a VOI that only covers the hippocampus to ease computational task of substructure segmentation. In order to preserve the spatial information of substructure, the coordinate the detected VOI is recorded during testing. Thus, the localization of ground truth hippocampus is used to supervise the localization model. To derive it, the manual contour is needed. For a set of MR images $I_{Img} \in R^{(w \times h \times d)}$, where w and h denote the width and height of the I_{Img}, d
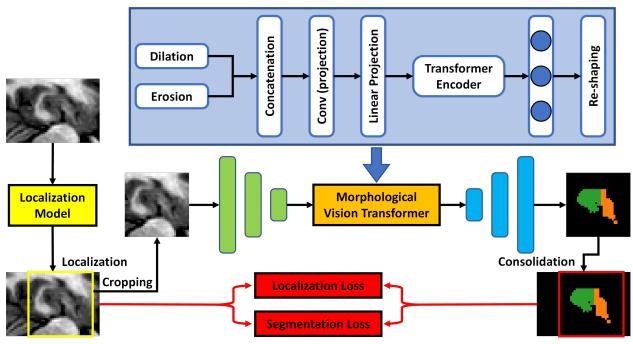
Figure 1. The workflow of the proposed morphological visual transformer learning-based hippocampus substructure segmentation.

represents its depth, and the corresponding physician-delineated hippocampus, $I_{Seg} = I_{seg}^p \cup I_{seg}^s I_{seg}^p$ denotes the hippocampus proper. I_{seg}^s denotes the parts of the subiculum. Based on the I_{Seg}, the bounding box that only covers the hippocampus can be derived. This bounding box is defined as the ground truth volumes-of-interest (VOI). The coordinate of the VOI is represented by $C = [x_c, y_c, z_c, w_c, h_c, d_c] \in R^6$, where x_c, y_c and z_c denote the center of hippocampus VOI, w_c, h_c and d_c denote the width, height and depth of the VOI along the 3D direction.

The localization model design is inspired by a recently developed focal modulation network, which is used in object detection.[26] The localization model includes a hierarchical contextualization, which is used for feature extraction from different hierarchical levels, a modulator, which combines the features from different levels, and a neural network layer works for location position estimation. The details of the localization model is explained as follows.

Given input MRI $I_{Img} \in R^{(w \times h \times d)}$, with a first convolution layer for feature map initiating F_0, a multi-scale hierarchy feature map set are collected via the steps defined as follows iteratively:

$$F_k = GeLu(Conv(F_{k-1})), \tag{1}$$

where F_{k-1} denotes the feature map from previous iteration, F_k is then derived by the operating convolution and Gaussian error linear units (GeLU) activation function.[13] After several iterations of Eq. (1), multi-hierarchical features are collected, we then match these feature maps to same size via interpolation and sum together

$$F_m = \Sigma_k BicubicInterpolate(F_k). \tag{2}$$

Then, by using a neural network layer, we aim to derive the estimation of C, labeled as $\hat{C} = [\hat{x}_c, \hat{y}_c, \hat{z}_c, \hat{w}_c, \hat{h}_c, \hat{d}_c]$, from the F_m. To achieve this aim, we set the loss function, as shown in Eq. (3) during the training of localization module.

$$L_{loc} = d((x_c, y_c, z_c), (\hat{x}_c, \hat{y}_c, \hat{z}_c)) + \lambda(\sqrt{w_c^2 - \hat{w}_c^2}/w + \sqrt{h_c^2 - \hat{h}_c^2}/h + \sqrt{d_c^2 - \hat{d}_c^2}/d), \tag{3}$$

where $d((x_c, y_c, z_c), (\hat{x}_c, \hat{y}_c, \hat{z}_c))$ denotes the Euclidean distance between the two centers (x_c, y_c, z_c) and $(\hat{x}_c, \hat{y}_c, \hat{z}_c)$.

2.3 Morphological visual transformer

For the next step, the MRI I_{Img} are cropped within a VOI box, whose center is defined as $\hat{C}$ This process mitigates the unrelated region for hippocampus segmentation and thus improve the efficiency of the model. To ensure the cropped image is uniformly sized for the following subnetwork, zero-padding is used. The processed image is then input into the morphological visual transformer (MVT). The MVT is built in an end-to-end fashion, meaning that the input and output share the same size. After several convolutional layers with a stride size of 2, the MVT uses two auto-learned morphological operators, dilation and erosion, to process the hidden feature maps. As compared to convolutional kernel with stride size of 2 or max-pooling layer, which can be regarded as a dilation with a flat square structuring element followed by a pooling, the learned morphological operator can be tuned to aggregate the most important information. This can further reduce the redundant and meaningless information for the next operator, the visual transformer, and therefore improve its performance. The output of the two morphological operators is then concatenated and fed into a projection convolutional layer and a linear projection operator to fit it to the input of visual transformer. A widely developed visual transformer is used.[24] Afterwards, several deconvolutional layers are applied until the output of this MVT model is equal in size to the input.

After the MVT step, consolidation can be used to transform the segmentation back to the original coordinate system (I_{img}), since the location information has been obtained from the localization model.

To supervise the MVT, a combination of two loss functions is used, which are generalized cross entropy loss L_{GCE} and generalized Dice loss L_{GD}. The L_{GCE} is used to evaluate the difference between the predicted label and the ground truth label at each voxel, which is defined as:

$$L_{GCE} = -\Sigma_i l_i \log \hat{l}_i \tag{4}$$

where l_i denotes the ground truth label at voxel i, $\hat{l}_i$ denotes the predicted label at voxel i.

The L_{GD} is used to address the issues about the voxel quantity imbalance of the segmented voxels (often a small portion of the whole image) and background (large portion), which is defined as:

$$L_{GD} = 1 - 2 \frac{\Sigma_i l_i \times \hat{l}_i + \epsilon}{\Sigma_i l_i^2 + \Sigma_i \hat{l}_i^2 + \epsilon} \tag{5}$$

where ϵ is a small value. The weighted sum of these two loss terms is then used to train the MVT model.

2.4 Dataset

In total, 260 T1w MR images from Medical Segmentation Decathlon were used in this study.[1] The Medical Segmentation Decathlon is a dataset consisting of T1-weighted magnetization-prepared rapid gradient echo (MPRAGE) MRIs of both healthy adults (ninety healthy adults) and adults with a non-affective psychotic disorder. The corresponding target Region of Interest (ROIs) were the anterior and posterior of the hippocampus, defined as the hippocampus proper and parts of the subiculum. This dataset was selected due to the precision needed to segment such a small object in the presence of a complex surrounding environment.

We conducted a five-fold cross-validation study on the first 200 T1w MR images. Then, a hold-out test was performed on the remaining 60 images using a model trained on the first 200 images. The segmentation was evaluated with multiple quantitative metrics including the Dice similarity coefficient (DSC), 95th percentile Hausdorff distance (HD95), mean surface distance (MSD), volume difference (VD) and center-of-mass distance (COMD). A Bland-Altman analysis and volumetric Pearson correlation analysis were also performed.

2.5 Implementation and evaluation

The investigated deep learning networks were designed using Python 3.6 and TensorFlow and implemented on a GeForce RTX 2080 GPU that had 12GB of memory. Optimization was performed using
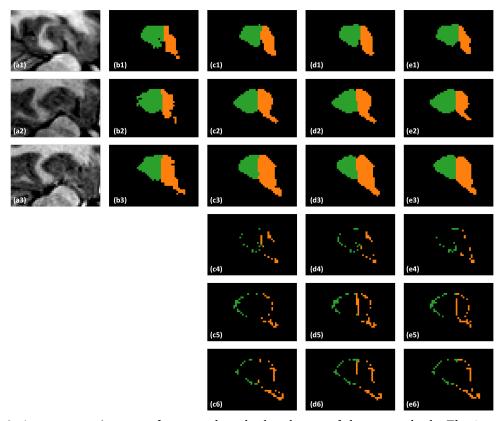
Figure 2. A representative case of proposed method and state-of-the-art methods. The 1st column shows MR images. The 2nd column shows the ground truth contour. The 3rd column shows the results of proposed method. The 4th column and 5th column show the results of cascaded U-Net and VIT, respectively. The last three rows are related to the absolute difference between segmented ones and ground truth ones.

the Adam gradient optimizer. The learning rate was 2×10-4. With the batch size setting of 20 during training, the percentage of utility of GPU memory is 96%. Once the network was trained, it only takes 1.5 mins for hippocampus segmentation. To demonstrate the utility of morphological operator, an ablation study was conducted. Namely, we tested the performance of the proposed method of with and without using morphological operator. To further demonstrate the significance of the proposed work, we compared the proposed method with another popular segmentation models, cascaded U-Net (CasU)[18] and visual transformer network (VIT).[24] Comparisons were performed using the same training and testing datasets and computational environment.

3 Results

3.1 Comparing with state-of-the-art

The visual comparison between the proposed method and comparing methods are shown in Fig. 2. As can be seen from the first row, the proposed method shows good agreement with the ground truth, whereas the comparing methods cannot. In the second row it is observed that misclassification of posterior part occurs for the cascaded U-Net. To better demonstrate the segmentation accuracy, we performed absolute subtraction of the segmentation results of the proposed method and comparing methods with the manual contour's binary masks. The difference images are shown in the fourth to sixth rows. As can be seen from the fifth and sixth rows, the difference images of the two comparing methods show greater error at the adjacent part between the hippocampus proper and parts of the subiculum.
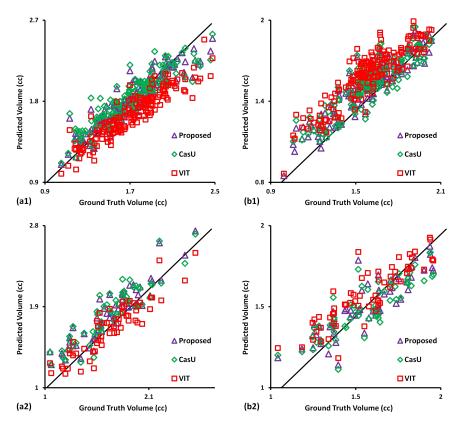
Figure 3. Bland-Altman analysis of the segmented volumes between ground truth (semi-log scale) against the proposed method and comparing methods. Each dot indicates a data point from the dataset for that model. (a) row denotes the results of five-fold cross-validation. (b) row denotes the results of hold-out test. First column denotes the segmentation of first substructure. Second column denotes the segmenting results of second substructure.

TABLE I. Numerical results (hippocampus proper) on 5-fold cross-validation of proposed method, cascaded U-Net and VIT, respectively.

	DSC	Jac	HD95 (mm)	MSD (mm)	RMSD (mm)	CMD (mm)
Proposed	0.900±0.029	0.819±0.047	1.156±0.277	0.426±0.115	0.688±0.122	0.491±0.300
CasU	0.891±0.031	0.804±0.049	1.329±0.478	0.466±0.123	0.744±0.156	0.581±0.363
VIT	0.894±0.027	0.809±0.044	1.195±0.296	0.441±0.107	0.706±0.112	0.573±0.297

The linear correlation coefficient calculated as target volume of ground truth and segmentation, is shown in Fig. 3. The linear correlation coefficient obtained using the proposed method was 0.999 and 0.993 on five-fold cross-validation and hold-out test, respectively. These values indicate a good agreement between the ground truth and proposed results, as compared to 0.989/0.983 and 0.991/0.979 obtained by the cascaded U-Net and VIT, respectively on five-fold cross-validation/hold-out test. On hold-out test, the VIT consistently underestimated the region , which became more pronounced for larger tumors.

The quantitative metrics of the proposed method and the alternate methods from the 200 cases' cross-validation and 60 cases' hold-out test are listed in Table 1 and 2, and Table 3 and 4, respectively. For the cross-validation experiment, the proposed model significantly outperformed Cascaded U-Net and VIT in all metrics. In five-fold cross-validation, the DSCs, HD95, MSD and CMD were 0.900±0.029 and 0.886±0.031, 1.156±0.277 and 1.133±0.264, 0.426±0.115 and 0.401±0.100, 0.491±0.300 and 0.738±0.452 for the hippocampus proper and parts of the subiculum, respectively.

In the hold-out test using external datasetthe proposed model is significantly superior to the alternate approaches, as shown in Table 3 and 4 in comparison with cascaded U-Net and VIT. In hold-out test, the DSCs, HD95, MSD and CMD were 0.881±0.033 and 0.863±0.034, 1.328±0.404 and 1.272±0.388, 0.494±0.113 and 0.466±0.112, 0.608±0.313 and 0.834±0.478 for the hippocampus proper and parts of

7

TABLE II. Numerical results (parts of the subiculum) on 5-fold cross-validation of proposed method, cascaded U-Net and VIT, respectively.

	DSC	Jac	HD95 (mm)	MSD (mm)	RMSD (mm)	CMD (mm)
Proposed	0.886±0.031	0.796±0.049	1.133±0.264	0.401±0.100	0.677±0.109	0.738±0.452
CasU	0.874±0.033	0.778±0.051	1.291±0.415	0.443±0.111	0.735±0.137	0.948±0.597
VIT	0.882±0.030	0.791±0.047	1.215±0.324	0.42±0.100	0.707±0.115	0.798±0.509

TABLE III. Numerical results (hippocampus proper) on hold-out test of proposed method, cascaded U-Net and VIT, respectively.

	DSC	Jac	HD95 (mm)	MSD (mm)	RMSD (mm)	CMD (mm)
Proposed	0.881±0.033	0.789±0.052	1.328±0.404	0.494±0.113	0.754±0.131	0.608±0.313
CasU	0.871±0.032	0.773±0.051	1.478±0.537	0.535±0.118	0.81±0.155	0.703±0.35
VIT	0.876±0.030	0.781±0.047	1.378±0.455	0.501±0.100	0.774±0.132	0.683±0.336

the subiculum, respectively. As compared to five-fold cross-validation, the hold-out test did slightly worse with slightly higher standard deviation, which may be caused by the training data's distribution not covering the range of cases in the hold-out test.

4 Discussion

A novel hippocampus segmentation method (called MVT) is proposed by introducing a localization mechanism to aid segmentation and designing the morphological visual transformer network for substructures segmentation. The localization model is used to detect the VOI of hippocampus. The end-to-end morphological vision transformer network is used to perform substructures segmentation within the hippocampus VOI. The substructures include the anterior and posterior regions of the hippocampus, which are defined as the hippocampus proper and parts of the subiculum. The vision transformer incorporates the dominant features extracted from MRI images and is improved by learning-based morphological operators. The morphological operators integrated into the vision transformer enhance the ability to separate the hippocampus structure into two substructures.

Due to limited computational resources, our method focused on domain incremental learning with a cropped region for analysis. We plan to test the performance of our method in a class incremental setup. As the visual transformer contains several orders of magnitude larger number of parameters due to the self-adapting process as compared to the traditional CNNs, it is essential to investigate an effective optimization method to reduce the amount of GPU memory allocation as well as simplify the overall ViT U-Net architecture.

Our MVT is a supervised method, which means it still requires accurate manual contours as training labels. Currently, there are semi-supervised learning methods that can learn features from unlabeled data. We will extend the proposed method with the ensemble approach to improve its generalization performance by integrating the supervision learning and semi-supervised learning methods from the limited labeled data and large-scale unlabeled data of MRIs in a future study.

The auto-segmentation of substructures of hippocampus has significant clinical relevance. For example, in hippocampal sparing whole brain radiation therapy (HA-WBRT),[22] current intensity modulated radiation treatment (IMRT) and arc-based VMAT techniques can reduce dose to the hippocampus

TABLE IV. Numerical results (parts of the subiculum) on hold-out test of proposed method, cascaded U-Net and VIT, respectively.

	DSC	Jac	HD95 (mm)	MSD (mm)	RMSD (mm)	CMD (mm)
Proposed	0.863±0.034	0.761±0.051	1.272±0.388	0.466±0.112	0.742±0.126	0.834±0.478
CasU	0.852±0.036	0.744±0.053	1.419±0.565	0.509±0.123	0.801±0.165	0.986±0.6
VIT	0.858±0.035	0.753±0.053	1.349±0.435	0.491±0.115	0.776±0.141	0.893±0.572

without sacrificing target coverage and homogeneity.[29] Further improvements in patient outcomes may be possible by considering substructures separately for optimal dose sparing; however, accurate segmentation is critical. With more accurate contouring of substructures of hippocampus, it is possible to have different dose constraints of these substructures in HA-WBRT,[21] allowing for better sparing of the critical part of the hippocampus.

5   Conclusion

We have developed a novel deep learning-based method to accurately segment the anterior and posterior of hippocampus. Our results showed good performance in terms of DSC and VD between the segmentation result and the ground truth.

Disclosures

The authors declare no conflicts of interest.

References

[1] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, AnnetteKopp-Schneider, Bennett A. Landman, Geert J. S. Litjens, Bjoern H. Menze, Olaf Ronneberger, Ronald M.Summers, Bram van Ginneken, Michel Bilello, Patrick Bilic, Patrick Ferdinand Christ, Richard K. G. Do, Marc J. Gollub, Stephan Heckers, Henkjan J. Huisman, William R. Jarnagin, Maureen McHugo, Sandy Napel, Jennifer S. Goli Pernicka, Kawal S. Rhode, Catalina Tobon-Gomez, Eugene Vorontsov, James Alastair Meakin, Sébastien Ourselin, Manuel Wiesenfarth, Pablo Arbeláez, Byeonguk Bae, Sihong Chen, Laura Alexandra Daza, Jian-Jun 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 Manuel Jorge Cardoso. The medical segmentation decathlon. *ArXiv*, abs/2106.05735, 2021.

[2] H. Braak and E. Braak. Neuropathological stageing of alzheimer-related changes. *Acta Neuropathol*, 82(4):239–59, 1991. ISSN 0001-6322 (Print) 0001-6322. doi: 10.1007/bf00308809. URL https://www.ncbi.nlm.nih.gov/pubmed/1759558.

[3] P. D. Brown, V. Gondi, S. Pugh, W. A. Tome, J. S. Wefel, T. S. Armstrong, J. A. Bovi, C. Robinson, A. Konski, D. Khuntia, D. Grosshans, T. L. S. Benzinger, D. Bruner, M. R. Gilbert, D. Roberge, V. Kundapur, K. Devisetty, S. Shah, K. Usuki, B. M. Anderson, B. Stea, H. Yoon, J. Li, N. N. Laack, T. J. Kruser, S. J. Chmura, W. Shi, S. Deshmukh, M. P. Mehta, and L. A. Kachnic. Hippocampal avoidance during whole-brain radiotherapy plus memantine for patients with brain metastases: Phase iii trial nrg oncology cc001. *J Clin Oncol*, 38(10):1019–1029, 2020. ISSN 0732-183X (Print) 0732-183x. doi: 10.1200/jco.19.02767. URL https://www.ncbi.nlm.nih.gov/pubmed/32058845.

[4] T. T. Brown. Individual differences in human brain development. *Wiley Interdiscip Rev Cogn Sci*, 8 (1-2), 2017. ISSN 1939-5078 (Print) 1939-5078. doi: 10.1002/wcs.1389. URL https://www.ncbi.nlm.nih.gov/pubmed/27906499.

[5] K. L. Canada, M. Botdorf, and T. Riggins. Longitudinal development of hippocampal subregions from early- to mid-childhood. *Hippocampus*, 30(10):1098–1111, 2020. ISSN 1050-9631 (Print) 1050-9631. doi: 10.1002/hipo.23218. URL https://www.ncbi.nlm.nih.gov/pubmed/32497411.

[6] Liang Cao, Long Li, Jifeng Zheng, Xin Fan, Feng Yin, Hui Shen, and Jun Zhang. Multi-task neural networks for joint hippocampus segmentation and clinical score regression. *Multimedia Tools and Applications*, 77(22):29669–29686, 2018. ISSN 1573-7721. doi: 10.1007/s11042-017-5581-1. URL https://doi.org/10.1007/s11042-017-5581-1.

[7] G. A. Carlesimo, F. Piras, M. D. Orfei, M. Iorio, C. Caltagirone, and G. Spalletta. Atrophy of presubiculum and subiculum is the earliest hippocampal anatomical marker of alzheimer's disease. *Alzheimers Dement (Amst)*, 1(1):24–32, 2015. ISSN 2352-8729 (Print) 2352-8729. doi: 10.1016/j.dadm.2014.12.001. URL https://www.ncbi.nlm.nih.gov/pubmed/27239489.

[8] T. Frodl, A. Schaub, S. Banac, M. Charypar, M. Jäger, P. Kümmler, R. Bottlender, T. Zetzsche, C. Born, G. Leinsinger, M. Reiser, H. J. Möller, and E. M. Meisenzahl. Reduced hippocampal volume correlates with executive dysfunctioning in major depression. *J Psychiatry Neurosci*, 31 (5):316–23, 2006. ISSN 1180-4882 (Print) 1180-4882. URL https://www.ncbi.nlm.nih.gov/pubmed/16951734.

[9] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang. A review of deep learning based methods for medical image multi-organ segmentation. *Phys Med*, 85:107–122, 2021. ISSN 1724-191X (Electronic) 1120-1797 (Linking). doi: 10.1016/j.ejmp.2021.05.003. URL https://www.ncbi.nlm.nih.gov/pubmed/33992856.

[10] J. W. Haller, G. E. Christensen, S. C. Joshi, J. W. Newcomer, M. I. Miller, J. G. Csernansky, and M. W. Vannier. Hippocampal mr imaging morphometry by means of general pattern matching. *Radiology*, 199(3):787–91, 1996. ISSN 0033-8419 (Print) 0033-8419. doi: 10.1148/radiology.199. 3.8638006. URL https://www.ncbi.nlm.nih.gov/pubmed/8638006.

[11] J. W. Haller, A. Banerjee, G. E. Christensen, M. Gado, S. Joshi, M. I. Miller, Y. Sheline, M. W. Vannier, and J. G. Csernansky. Three-dimensional hippocampal mr morphometry with high-dimensional transformation of a neuroanatomic atlas. *Radiology*, 202(2):504–510, 1997. ISSN 0033-8419. doi: 10.1148/radiology.202.2.9015081. URL https://doi.org/10.1148/radiology.202.2.9015081.

[12] Y. Hao, T. Wang, X. Zhang, Y. Duan, C. Yu, T. Jiang, and Y. Fan. Local label learning (lll) for subcortical structure segmentation: application to hippocampus segmentation. *Hum Brain Mapp*, 35(6):2674–97, 2014. ISSN 1065-9471 (Print) 1065-9471. doi: 10.1002/hbm.22359. URL https://www.ncbi.nlm.nih.gov/pubmed/24151008.

[13] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv*, arXiv:1606.08415, 2016. URL https://doi.org/10.48550/arXiv.1606.08415.

[14] Kourosh Jafari-Khouzani, Kost V. Elisevich, Suresh Patel, and Hamid Soltanian-Zadeh. Dataset of magnetic resonance images of nonepileptic subjects and temporal lobe epilepsy patients for validation of hippocampal segmentation techniques. *Neuroinformatics*, 9:335–346, 2010.

[15] Y. Lei, S. Tian, X. He, T. Wang, B. Wang, P. Patel, A. B. Jani, H. Mao, W. J. Curran, T. Liu, and X. Yang. Ultrasound prostate segmentation based on multidirectional deeply supervised v-net. *Med Phys*, 46(7):3194–3206, 2019. ISSN 2473-4209 (Electronic) 0094-2405 (Linking). doi: 10. 1002/mp.13577. URL https://www.ncbi.nlm.nih.gov/pubmed/31074513.

[16] Yang Lei, Yabo Fu, Tonghe Wang, Richard L. J. Qiu, Walter J. Curran, Tian Liu, and Xiaofeng Yang. *Deep Learning Architecture Design for Multi-Organ Segmentation*, volume Chapter 7, Part 2 of *Auto-Segmentation for Radiation Oncology: State of the Art*. CRC Press., Boca Raton, 1st edition edition, 2021. ISBN 9780429323782. doi: 10.1201/9780429323782-9. URL https://www.taylorfrancis.com/chapters/edit/10.1201/9780429323782-9/deep-learning-architecture-design-multi-organ-segmentation-yang-lei-yabo-fu-tonghe-wang-richar context=ubx&refId=0470c1fe-34c9-4c46-b608-4a1963c64be0.

[17] M. Lin, J. F. Wynne, B. Zhou, T. Wang, Y. Lei, W. J. Curran, T. Liu, and X. Yang. Artificial intelligence in tumor subregion analysis based on medical imaging: A review. *J Appl Clin Med Phys*, 22(7):10–26, 2021. ISSN 1526-9914 (Electronic) 1526-9914 (Linking). doi: 10.1002/acm2.13321. URL https://www.ncbi.nlm.nih.gov/pubmed/34164913.

[18] Hongying Liu, Xiongjie Shen, Fanhua Shang, Feihang Ge, and Fei Wang. Cu-net: Cascaded u-net with loss weighted sampling for brain tumor segmentation. In Dajiang Zhu, Jingwen Yan, Heng Huang, Li Shen, Paul M. Thompson, Carl-Fredrik Westin, Xavier Pennec, Sarang Joshi, Mads Nielsen, Tom Fletcher, Stanley Durrleman, and Stefan Sommer, editors, *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy*, pages 102–111. Springer International Publishing. ISBN 978-3-030-33226-6.

[19] T. L. McHugh, A. J. Saykin, H. A. Wishart, L. A. Flashman, H. B. Cleavinger, L. A. Rabin, A. C. Mamourian, and L. Shen. Hippocampal volume and shape analysis in an older adult population. *Clin Neuropsychol*, 21(1):130–45, 2007. ISSN 1385-4046 (Print) 1385-4046. doi: 10.1080/13854040601064534. URL https://www.ncbi.nlm.nih.gov/pubmed/17366281.

[20] S. Nobakht, M. Schaeffer, N. D. Forkert, S. Nestor, E. Black S, P. Barber, and Initiative The Alzheimer's Disease Neuroimaging. Combined atlas and convolutional neural network-based segmentation of the hippocampus from mri according to the adni harmonized protocol. *Sensors (Basel)*, 21(7), 2021. ISSN 1424-8220. doi: 10.3390/s21072427. URL https://www.ncbi.nlm.nih.gov/pubmed/33915960.

[21] I. Popp, A. L. Grosu, J. T. Fennell, M. Fischer, D. Baltas, and R. Wiehle. Optimization of hippocampus sparing during whole brain radiation therapy with simultaneous integrated boost-tutorial and efficacy of complete directional hippocampal blocking. *Strahlenther Onkol*, 198(6):537–546, 2022. ISSN 0179-7158 (Print) 0179-7158. doi: 10.1007/s00066-022-01916-3.

[22] Ilinca Popp, Stephan Rau, Mandy Hintz, Julius Schneider, Angelika Bilger, Jamina Tara Fennell, Dieter Henrik Heiland, Thomas Rothe, Karl Egger, Carsten Nieder, Horst Urbach, and Anca Ligia Grosu. Hippocampus-avoidance whole-brain radiation therapy with a simultaneous integrated boost for multiple brain metastases. *Cancer*, 126(11):2694–2703, 2020. ISSN 0008-543X. doi: https://doi.org/10.1002/cncr.32787. URL https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.1002/cncr.32787.

[23] Q. Qiu, Z. Yang, S. Wu, D. Qian, J. Wei, G. Gong, L. Wang, and Y. Yin. Automatic segmentation of hippocampus in hippocampal sparing whole brain radiotherapy: A multitask edge-aware learning. *Med Phys*, 48(4):1771–1780, 2021. ISSN 0094-2405. doi: 10.1002/mp.14760. URL https://www.ncbi.nlm.nih.gov/pubmed/33555048.

[24] Amin Ranem, Camila Gonzalez, and Anirban Mukhopadhyay. Continual hippocampus segmentation with transformers. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3710–3719. doi: doi:10.1109/CVPRW56347.2022.00415.

[25] D. H. Salat, J. J. Chen, A. J. van der Kouwe, D. N. Greve, B. Fischl, and H. D. Rosas. Hippocampal degeneration is associated with temporal and limbic gray matter/white matter tissue contrast in alzheimer's disease. *Neuroimage*, 54(3):1795–802, 2011. ISSN 1053-8119 (Print) 1053-8119. doi: 10.1016/j.neuroimage.2010.10.034. URL https://www.ncbi.nlm.nih.gov/pubmed/20965261.

[26] M. R. Siadat, H. Soltanian-Zadeh, and K. V. Elisevich. Knowledge-based localization of hippocampus in human brain mri. *Comput Biol Med*, 37(9):1342–60, 2007. ISSN 0010-4825 (Print) 0010-4825. doi: 10.1016/j.compbiomed.2006.12.010. URL https://www.ncbi.nlm.nih.gov/pubmed/17339035.

[27] L. A. van de Pol, A. Hensel, W. M. van der Flier, P. J. Visser, Y. A. Pijnenburg, F. Barkhof, H. J. Gertz, and P. Scheltens. Hippocampal atrophy on mri in frontotemporal lobar degeneration and alzheimer's disease. *J Neurol Neurosurg Psychiatry*, 77(4):439–42, 2006. ISSN 0022-3050 (Print) 0022-3050. doi: 10.1136/jnnp.2005.075341. URL https://www.ncbi.nlm.nih.gov/pubmed/16306153.

[28] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich. Multi-atlas segmentation with joint label fusion. *IEEE Trans Pattern Anal Mach Intell*, 35(3):611–23, 2013. ISSN 0162-8828 (Print) 0098-5589. doi: 10.1109/tpami.2012.143. URL https://www.ncbi.nlm.nih.gov/pubmed/22732662.

[29] A. H. L. Yuen, P. M. Wu, A. K. L. Li, and P. C. Y. Mak. Volumetric modulated arc therapy (vmat) for hippocampal-avoidance whole brain radiation therapy: planning comparison with dual-arc and split-arc partial-field techniques. *Radiat Oncol*, 15(1):42, 2020. ISSN 1748-717x. doi: 10.1186/s13014-020-01488-5.