CT-Based Hippocampus Segmentation with Dual-Decoder Network (HDD-Net)

Wonjun Son[1], Ji Young Lee[2], Sung Jun Ahn[3], Hyunyeol Lee[1*]

[1]School of Electronic and Electrical Engineering & IEDT,
Kyungpook National University, Daegu, Republic of Korea
[2]Department of Radiology, Seoul St. Mary's Hospital,
Catholic University College of Medicine, Seoul, Republic of Korea
[3]Department of Radiology, Gangnam Severance Hospital,
Yonsei University College of Medicine, Seoul, Republic of Korea
*Corresponding author: hyunyeollee@knu.ac.kr

Abstract. The hippocampus in the brain performs a pivotal role for memory formation, spatial navigation, and emotional regulation. Its volume and morphology are known to change with the progression of neurodegenerative diseases such as Alzheimer's disease. Hence, hippocampal atrophy serves as a key biomarker for early diagnosis and monitoring of such diseases. Whereas MRI has been predominantly employed in that regard due to its excellent soft-tissue contrast, CT-based segmentation of the structure has been relatively far less explored because the modality results in ambiguous boundaries between brain subregions. This study aims to address this technical challenge, achieving accurate segmentation of the hippocampus on CT images. To this end, we develop a deep learning model, termed 'Hippocampus Dual Decoder Network (HDD-Net)', characterized by the following four major components: 1) parallel, dual decoders that segment the hippocampal region and its boundaries, respectively, 2) a single, shared encoder in which features combined across multiple blocks are refined via attention, 3) a feature fusion module (FFM) that performs inter-decoder featural supplements, and 4) a cross loss to jointly optimize segmentation and edge predictions. HDD-Net was validated using both internal and external datasets, with its performance assessed using Dice similarity coefficient (DSC) and intersection-over-union (IoU). Our model yielded DSC = 0.823 ± 0.03 and IoU = 0.701 ± 0.04, and DSC = 0.759 ± 0.07 and IoU = 0.617 ± 0.09 for internal and external test datasets, respectively, outperforming seven other SOTA methods. Furthermore, volumetric analysis revealed a good agreement between MRI- and CT-derived hippocampal masks. Our findings suggest feasibility of CT-based hippocampal segmentation via HDD-Net, as a cost-effective alternative to MRI. The implementation of HDD-Net is available at https://github.com/sonwonjun103/HDD_Net.

Keywords: Computed Tomography, Hippocampus Segmentation, Hippocampus Dual Decoder Network (HDD-Net), Deep Learning

1 Introduction

The hippocampus in the brain plays a crucial role in memory formation, spatial navigation, and emotional regulation, and abnormalities in its volume and morphology linked to Alzheimer's disease, epilepsy, PTSD, schizophrenia, and depression [1-3]. Magnetic resonance imaging (MRI) has been regarded as the gold standard for hippocampal segmentation due to its excellent soft-tissue contrast. Nevertheless, high cost and long acquisition time make its application difficult in certain clinical settings, particularly with limited-resource environments. Computed tomography (CT) can be considered as a more accessible and cost-effective alternative to MRI. However, near-flat contrast across brain subregions in CT images renders hippocampal segmentation from the modality challenging.

Recent advances in CNN-based architectures like U-Net [4] and transformer-based models such as UNETR [5] and TransUNet [6] have significantly improved segmentation performance. Accordingly, deep learning (DL)-based medical image segmentation has been extensively explored [7, 8], yet a majority of which are on MRI [9-12]. One of very few studies attempting DL segmentation of the hippocampus from CT scans is AG-3D ResNet by Portal et al. [13]. The authors integrated attention mechanisms into a 3D ResNet [14] backbone as a means to enhance feature extraction, and showed feasibility of CT-based hippocampal segmentation using DL.

In this work, we were aimed to enhance performance of DL segmentation of hippocampus on CT head images. To achieve this goal, we conceived a novel DL model, termed "Hippocampus Dual Decoder Network (HDD-Net)", and evaluated its performance in reference to MRI-derived hippocampal labels.

2 Methods

The overall architecture of HDD-Net, shown in Fig. 1, is designed to address the challenges of hippocampal segmentation on CT images. It consists of four key components: dual-decoder, a shared encoder, feature fusion module (FFM), and a cross loss function. By leveraging its dual-decoder structure, the network takes preprocessed volumetric 3D CT images as input and produces two outputs: hippocampal masks and corresponding edges. Details in each component are described in the following subsections.

2.1 Dual Decoder

The dual decoder architecture enables HDD-Net to leverage both volume- and edge-specific information, thereby achieving performance elevation in hippocampal segmentation. It consists of a Segmentation Decoder, which predicts the hippocampal region, and an Edge Decoder, which extracts boundaries of the structure. It ensures that the model captures complementary aspects of the hippocampal structure, addressing the challenges posed by its small size and complex anatomy.

Both decoders share a same structure with different training parameters, comprising a series of up-sampling and convolutional layers to progressively restore spatial resolution. Each decoder block contains two convolutional layers with kernel size $3 \times 3 \times 3$ followed by batch normalization and ReLU activation with an up-sampling layer applied at the end of each block. The filter sizes in the decoder layers progressively decrease, following a sequence of (512, 256, 128, 64), with a final kernel size $1 \times 1 \times 1$ convolution layer at the end that uses a reduced filter size of 32 and performs the convolution operation only once. A soft-max function is applied at the end of each decoder to generate normalized probability maps.

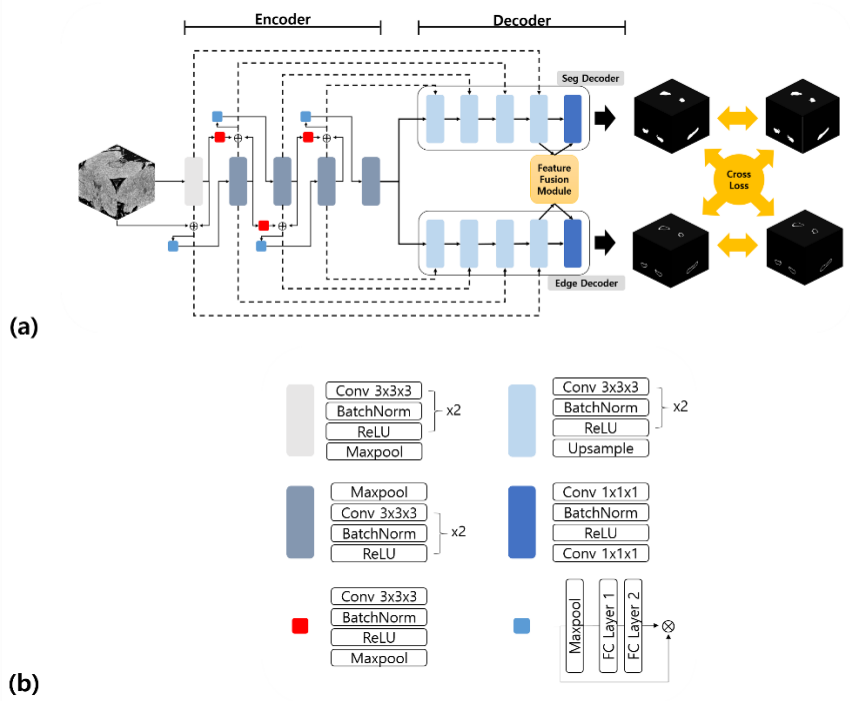

Fig. 1. The architecture of the proposed HDD-Net. (a) The network consists of dual decoders (Segmentation Decoder and Edge Decoder), a shared encoder, a feature fusion module (FFM), and a cross loss function. (b) Detailed structures of the convolutional blocks used in the encoder and decoders.

2.2 Shared Encoder

The shared encoder is responsible for extracting multi-scale spatial and contextual features from the input volumetric 3D CT images. It consists of five convolutional blocks with filter sizes set to (32, 64, 128, 256, 512).

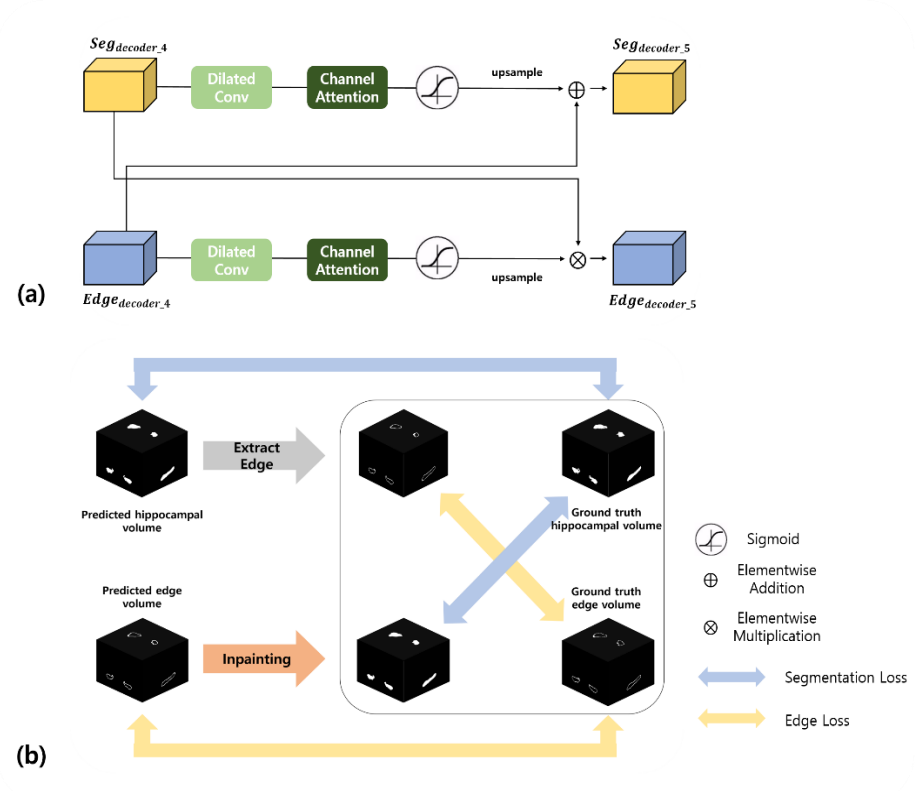

Fig. 2. (a) The feature fusion module (FFM) enables interaction between the Segmentation Decoder and Edge Decoder through dilated convolution, channel attention, and feature integration via elementwise operations. (b) The cross loss jointly optimizes volumetric and edge predictions by aligning them with the respective ground truths using segmentation and edge loss calculations.

The first block begins by processing the input using an initialization block (Init), which applies two $3 \times 3 \times 3$ convolution layers, each followed by batch normalization, ReLU activation, and max-pooling layer, produced the first-level feature map (E_1). From the second block onward, each block (O_{conv}) concatenates features from the outputs of the two preceding blocks after being downsampled (O_{reduce}) to match dimensions. A channel attention (O_{CA}) [15] is used to emphasize the most relevant channels in the concatenated features, improving feature quality. These refined features are then passed through max-pooling layer followed by two convolution layers with batch normalization and ReLU activation. Additionally, skip connections from each block to corresponding decoder blocks help retain spatial information during reconstruction.

Formally, the process can be described as:

$$E_1 = \text{Init}(\text{Input})$$
$$E_2 = O_{conv}(O_{CA}(\text{Concat}(\text{Input}, E_1)))$$

$$E_3=O_{conv}(O_{CA}(Concat(O_{reduce}(E_1),E_2)))$$
$$E_4=O_{conv}(O_{CA}(Concat(O_{reduce}(E_2),E_3)))$$
$$E_5=O_{conv}(O_{CA}(Concat(O_{reduce}(E_3),E_4)))$$

where $E_i(i = 1,2,3,4,5)$ represents the feature map from each ith blocks.

2.3 Feature Fusion Module (FFM)

FFM enables interaction between the Segmentation Decoder and the Edge Decoder, as illustrated in Fig 2(a). By combining the features extracted by the two decoders, the FFM provides a comprehensive representation that incorporates both volumetric and edge-specific information, leading to more precise segmentation results.

The fusion process begins with a dilated convolution [16] with kernel size $3 \times 3 \times 3$ (dilation =1, 3, 5, 7), which expands the receptive field to capture multi-scale contextual information. This is followed by a channel attention [15] that dynamically recalibrates feature maps to prioritize the most relevant channels. To seamlessly integrate features between the decoders, the FFM performs targeted operations that enhance both volumetric and edge predictions. Specifically, the feature map from the Segmentation Decoder is combined with the FFM-processed edge feature map through elementwise multiplication, enabling the model to enhance edge predictions by leveraging volumetric features. Simultaneously, the feature map from the Edge Decoder is combined with the FFM-processed segmentation feature map using elementwise addition, allowing the model to refine volumetric predictions by incorporating detailed edge information.

This bidirectional flow in information ensures effective collaboration between the decoders, enabling model to capture both global structural details and fine-grained edges with high precision.

2.4 Cross Loss

The cross loss, depicted in Figure 2(b), is a custom-designed function that jointly optimize the outputs of the Segmentation Decoder and the Edge Decoder, ensuring accurate prediction of both hippocampal volumes and edges. By integrating segmentation loss (L_{seg}) and edge loss (L_{edge}), the cross loss enables the model to focus on complementary aspects of segmentation accuracy. This dual optimization is achieved through a combination of dice loss and cross-entropy loss ($L_{dice\&CE}$), which together enhance both overlap-based and pixel-wise classification performance.

The segmentation loss measures the difference between the predicted hippocampal volume (Y_{P_V}) and the ground truth hippocampal volume (Y_{GT_V}). Additionally, it evaluates the consistency of in-painted predicted edge volume ($Y_{P_{E\to P_V}'}$) with the ground truth hippocampal volume (Y_{GT_V}), ensuring alignment between volumetric and boundary features:

$$L_{seg} = L_{dice\ \&\ CE}\left(Y_{P_V}, Y_{gt_V}\right) + L_{dice\ \&\ CE}\left(Y_{P_{E\to P_V}'}, Y_{gt_V}\right) \quad (1)$$

The edge loss ensures the consistency of predicted edge volumes (Y_{P_E}) with the ground truth edge volume (Y_{GT_E}) and evaluates edges extracted from predicted volumes ($Y_{P_{V \to P_E}}$):

$$L_{edge} = L_{dice\ \&\ CE}(Y_{P_E}, Y_{gt_E}) + L_{dice\ \&\ CE}\left(Y_{P_{V \to P_E}}, Y_{gt_E}\right) \tag{2}$$

Combining Eq. (1) and Eq. (2), the cross loss (L_{cross}) is defined as:

$$L_{cross} = \alpha L_{seg} + \beta L_{edge} \tag{3}$$

where α and β are weighting factors. In this work, we set $\alpha = 1$ and $\beta = 1$.

3 Experiments and Results

3.1 Experiments

Datasets. We collected datasets from Gangnam Severance Hospital (GSH) and Seoul St. Mary's Hospital (SSMH) with approval of both institutional review boards. Informed consent from the patients was waived due to the study's retrospective nature. A total of 150 neurologically healthy individuals from GSH who underwent both brain CT and T1-weighted MRI [17] within three months were included, meeting specific imaging criteria. The GSH dataset was split into a training set (n=120) and internal test set (n=30), while an external validation set (n=47) from SSMH was selected using the same criteria.

Data preprocessing. CT and MRI images were converted from DICOM to NIFTI for efficient handling and processing of 3D images. CT images were registered to MRI using SPM12 [18], normalized to HU range (-20 to 100), and min-max scaled to [0, 1], MR images were processed with Freesurfer [19], yielding hippocampal masks as ground truth, from which edges were extracted. Finally, co-registered CT-MRI pairs along with ground-truth labels were center-cropped from (256, 256, 256) to (96, 128, 128).

Implementation details. The proposed HDD-Net model implemented in this work comprises 100,195,112 learnable parameters, which were randomly initialized at the start of training. Hyperparameters we used during the training phase included a batch size of 4, a total of 150 epochs, and a learning rate of 0.0001 with the Adam optimizer. The network was implemented and trained using PyTorch 2.3.1 on two NVIDIA GeForce RTX 3090 Ti GPUs, each equipped with 24GB of memory. The implementation of HDD-Net is available at https://github.com/sonwonjun103/HDD_Net.

Evaluation. The performance of the proposed model was evaluated using two standard metrics: Dice similarity coefficient (DSC) and Intersection over Union (IoU).

Additionally, Bland-Altman analysis was conducted to assess volumetric agreement between MRI-based ground truth and CT-based segmentation results.

3.2 Results

Qualitative analysis. Fig 3 illustrates representative segmentation results for test subject. The first column shows CT image, while the second column shows the corresponding MR image. The final column shows the overlaid the ground truth hippocampal segmentation with the prediction map generated by HDD-Net. Additionally, the 3D renderings further validate the model's accuracy, as the predicted hippocampal segmentation maps closely resemble the hippocampus derived from MR images. This consistency across modalities emphasizes the models' capability to address the challenges of hippocampus segmentation in non-contrast CT images, producing results comparable to MRI-based ground truths.

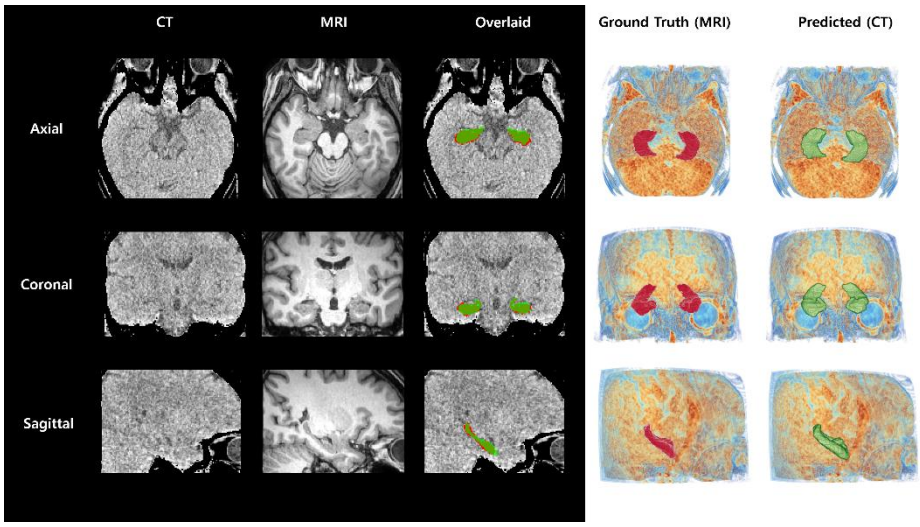

Fig. 3. Visualization of hippocampus segmentation results. The left section shows axial, coronal, and sagittal views of CT images, MRI ground truth, and overlaid segmentation results. (Red: ground truth volume, Green: overlaid volume, Light green: predicted volume). The right section provides 3D renderings of hippocampal volumes from the MRI ground truth and the predicted volume by the proposed model.

Evaluation scores. On internal datasets, the model achieved a mean DSC of 0.823 ± 0.03 and IoU of 0.701 ± 0.05. For external datasets, the model achieved a mean DSC of 0.759 ± 0.07 and IoU of 0.617 ± 0.09. To evaluate the performance of HDD-Net, a comparative study was conducted against widely-used U-Net based models in medical segmentation, summarized in Table 1. These models include the standard U-Net, it enhanced version such as Attention U-Net [21] and nnU-Net [22], as well as models integrating transformer blocks like 3D TransU-Net [6], Swin U-NetR [23], and U-NetR

[5]. Additionally, the comparison included the AG-3D ResNet proposed by Portal et al [13]. Our model outperformed these models across both internal and external datasets.

Table 1. Performance comparison of HDD-Net with U-Net based segmentation models.

Models	Internal (n=30)		External (n=47)	
	DSC	IoU	DSC	IoU
U-Net [20]	0.784±0.04	0.647±0.06	0.718±0.07	0.565±0.08
Attention U-Net [21]	0.775±0.04	0.635±0.06	0.694±0.08	0.538±0.09
nnUNet [22]	0.769±0.03	0.626±0.05	0.588±0.06	0.419±0.06
TransUNet [6]	0.773±0.03	0.632±0.05	0.589±0.05	0.420±0.05
UNETR [5]	0.656±0.06	0.491±0.06	0.508±0.13	0.351±0.11
Swin UNETR [23]	0.650±0.07	0.485±0.07	0.558±0.14	0.400±0.13
AG-3D ResNet [13]	0.762±0.06	0.620±0.07	0.673±0.13	0.516±0.11
HDD-Net (Ours)	**0.823±0.03**	**0.701±0.05**	**0.759±0.07**	**0.617±0.09**

Bland-Altman analysis. To evaluate the agreement between MRI-based and DL-based CT segmentations, bland-altman analysis conducted for internal and external datasets. For the overall volumetric differences, it showed a mean difference $0.53cm^3$ /$1.17cm^3$ for internal and external datasets, respectively (Fig 4(a) and Fig 4(b)), indicating good agreement between the two hippocampal volumes.

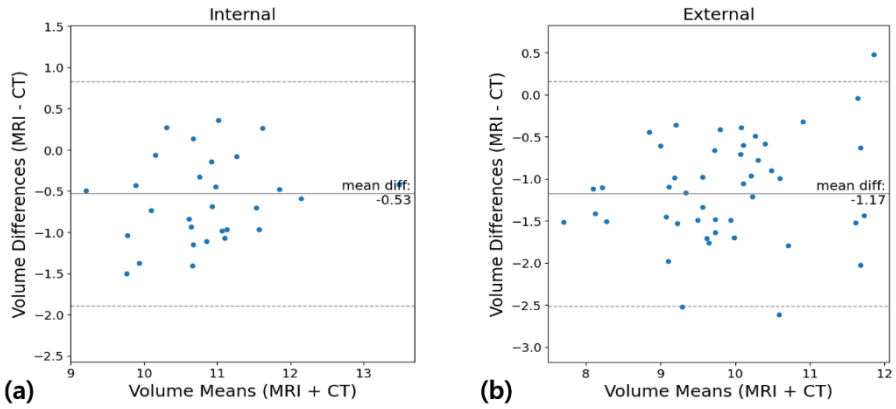

Fig. 4. Bland-Altman plots illustrating the differences in hippocampal volumetric measurements between MRI-based segmentation and DL-based segmentation on CT scans for (a) internal datasets (n=30) and (b) external datasets (n=47).

4 Discussion and conclusions

HDD-Net is a novel dual-decoder deep learning model designed for hippocampal segmentation from CT images, addressing the limitations of MRI dependency in clinical applications. By integrating a FFM for bidirectional feature exchange and a cross loss for joint optimization, HDD-Net achieves superior segmentation performance. The

model outperforms existing U-Net and transformer-based methods on both internal and external datasets, demonstrating its effectiveness in CT-based hippocampal segmentation. Notably, this method enables hippocampal volumetric analysis in environments where MRI is limited or unavailable, broadening its clinical applicability.

In conclusion, HDD-Net presents a promising approach to hippocampal segmentation from CT, expanding accessibility to hippocampal volumetric analysis in MRI-limited environments. By eliminating the need for MRI in segmentation tasks, this method could enhance clinical workflows in Alzheimer's disease assessment and hippocampal avoidance whole-brain radiotherapy.

Acknowledgments. This work was supported by the IITP-Innovative Human Resource Development for Local Intellectualization program grant (IITP-2025-RS-2022-00156389, 50%), and the Commercialization Promotion Agency for R&D Outcomes (COMPA) grant (RS-2023-00304695, 50%), both funded by the Korean Government (Ministry of Science and ICT).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

5 References

1. Dubois, B., Feldman, H.H., Jacova, C., DeKosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G.: Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. The Lancet Neurology 6, 734-746 (2007)
2. Bonne, O., Brandes, D., Gilboa, A., Gomori, J.M., Shenton, M.E., Pitman, R.K., Shalev, A.Y.: Longitudinal MRI study of hippocampal volume in trauma survivors with PTSD. American Journal of Psychiatry 158, 1248-1251 (2001)
3. Bremner, J.D., Narayan, M., Anderson, E.R., Staib, L.H., Miller, H.L., Charney, D.S.: Hippocampal volume reduction in major depression. American Journal of Psychiatry 157, 115-118 (2000)
4. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, pp. 234-241. Springer, (2015)
5. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 574-584. (2022)
6. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
7. Son, W.J., Ahn, S.J., Lee, J.Y., Lee, H.: Automated Brain Segmentation on Computed Tomographic Images Using Perceptual Loss Based Convolutional Neural Networks. Investigative Magnetic Resonance Imaging 28, 193-201 (2024)

8. Kim, J., Lee, H., Oh, S.S., Jang, J., Lee, H.: Automated Quantification of Total Cerebral Blood Flow from Phase-Contrast MRI and Deep Learning. Journal of Imaging Informatics in Medicine 37, 563-574 (2024)
9. Ataloglou, D., Dimou, A., Zarpalas, D., Daras, P.: Fast and precise hippocampus segmentation through deep convolutional neural network ensembles and transfer learning. Neuroinformatics 17, 563-582 (2019)
10. Zeng, D., Li, Q., Ma, B., Li, S.: Hippocampus segmentation for preterm and aging brains using 3D densely connected fully convolutional networks. IEEE Access 8, 97032-97044 (2020)
11. Chen, X., Peng, Y., Li, D., Sun, J.: DMCA-GAN: Dual Multilevel Constrained Attention GAN for MRI-Based Hippocampus Segmentation. Journal of Digital Imaging 36, 2532-2553 (2023)
12. Xiao, Z., Zhang, Y., Deng, Z., Liu, F.: Light3DHS: A lightweight 3D hippocampus segmentation method using multiscale convolution attention and vision transformer. NeuroImage 292, 120608 (2024)
13. Porter, E., Fuentes, P., Siddiqui, Z., Thompson, A., Levitin, R., Solis, D., Myziuk, N., Guerrero, T.: Hippocampus segmentation on noncontrast CT using deep learning. Medical physics 47, 2950-2961 (2020)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. (2015)
15. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), pp. 3-19. (2018)
16. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 (2015)
17. Mugler III, J.P., Brookeman, J.R.: Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). Magnetic resonance in medicine 15, 152-157 (1990)
18. Flandin, G., Friston, K.J.: Statistical parametric mapping (SPM). Scholarpedia 3, (2008)
19. Fischl, B.: FreeSurfer. Neuroimage 62, 774-781 (2012)
20. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19, pp. 424-432. Springer (2016)
21. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
22. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods 18, 203-211 (2021)

23. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop, pp. 272-284. Springer (2022)