



Automatic brain structure segmentation for ^{18}F -fluorodeoxyglucose positron emission tomography/magnetic resonance images via deep learning

Zhenxing Huang^{1#}, Han Liu^{1#}, Yaping Wu^{2#}, Wenbo Li¹, Jun Liu³, Ruodai Wu⁴, Jianmin Yuan⁵, Qiang He⁵, Zhe Wang⁵, Ke Zhang⁶, Dong Liang¹, Zhanli Hu¹, Meiyun Wang², Na Zhang¹

¹Lauterbur Research Center for Biomedical Imaging, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China; ²Department of Medical Imaging, Henan Provincial People's Hospital & People's Hospital of Zhengzhou University, Zhengzhou, China; ³Department of Radiology, The Second Xiangya Hospital, Central South University, Changsha, China; ⁴Department of Radiology, Shenzhen University General Hospital, Shenzhen University Clinical Medical Academy, Shenzhen, China; ⁵Central Research Institute, United Imaging Healthcare Group, Shanghai, China; ⁶Department of Diagnostic and Interventional Radiology, Heidelberg University Hospital, Heidelberg, Germany

Contributions: (I) Conception and design: Z Huang, H Liu, M Wang, N Zhang; (II) Administrative support: D Liang, Z Hu, M Wang; (III) Provision of study materials or patients: J Liu, R Wu, Y Wu; (IV) Collection and assembly of data: J Yuan, Q He, Z Wang; (V) Data analysis and interpretation: M Wang, N Zhang, W Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work and should be considered as co-first authors.

Correspondence to: Na Zhang, PhD. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Shenzhen University Town, Shenzhen 518055, China. Email: na.zhang@siat.ac.cn; Meiyun Wang, MD. Department of Medical Imaging, Henan Provincial People's Hospital & People's Hospital of Zhengzhou University, No. 7 Weiwu, Zhengzhou 450003, China. Email: mywang@zzu.edu.cn.

Background: Brain structure segmentation is of great value in diagnosing brain disorders, allowing radiologists to quickly acquire regions of interest and assist in subsequent analyses, diagnoses and treatment. Current brain structure segmentation methods are usually applied to magnetic resonance (MR) images, which provide higher soft tissue contrast and better spatial resolution. However, fewer segmentation methods are conducted on a positron emission tomography/magnetic resonance imaging (PET/MRI) system that combines functional and structural information to improve analysis accuracy.

Methods: In this paper, we explore a dual-modality image segmentation model to segment brain ^{18}F -fluorodeoxyglucose (^{18}F -FDG) PET/MR images based on the U-Net architecture. This model takes registered PET and MR images as parallel inputs, and four evaluation metrics (Dice score, Jaccard coefficient, precision and sensitivity) are used to evaluate segmentation performance. Moreover, we also compared the proposed approach with other single-modality segmentation strategies, including PET-only segmentation and MRI-only segmentation.

Results: The experiments were conducted on the clinical head data of 120 patients, and the results show that the proposed algorithm accurately delineates brain volumes of interest (VOIs), achieving superior performance with $84.24\% \pm 1.44\%$ Dice score, $74.36\% \pm 2.40\%$ Jaccard, $84.33\% \pm 1.56\%$ precision and $84.73\% \pm 1.56\%$ sensitivity. Furthermore, compared with directly using the FreeSurfer toolkit, the proposed method reduced the segmentation time, which only needs 20 seconds to segment the whole brain for each patient.

Conclusions: We present a deep learning-based method for the joint segmentation of anatomical and functional PET/MR images. Compared with other single-modality methods, our method greatly improved the accuracy of brain structure delineation, which shows great potential for brain analysis.

Keywords: Positron emission tomography/magnetic resonance (PET/MR); automatic segmentation; brain structure; deep learning

Submitted Oct 13, 2022. Accepted for publication Apr 20, 2023. Published online Jun 08, 2023.

doi: 10.21037/qims-22-1114

View this article at: <https://dx.doi.org/10.21037/qims-22-1114>

Introduction

Brain scans can be applied to detect signs of various brain diseases, such as dementia, Parkinson's disease (PD) and Alzheimer's disease (AD) (1-3), and segmentation of the brain facilitates structural localization and morphological feature extraction, as well as the identification of diagnostic biomarkers. Magnetic resonance imaging (MRI) plays an important role in the study of the human brain because of its good performance in revealing brain anatomy, pathology and function (4,5). Although magnetic resonance (MR) images process higher soft tissue contrast and better spatial resolution, positron emission tomography (PET) images, which can effectively detect early lesions based on the metabolism of different tracers in different areas (6-8), are still needed to assist in the localization, visualization, and assessment of abnormal areas (9,10). However, most of the current segmentation methods are mainly based on MRI, and few studies (11-13) have focused on the PET/MRI dual modality, ignoring the improvement in segmentation accuracy by utilizing both functional and structural information.

Medical image segmentation is expected to divide the image into several regions that preferably correspond to anatomical regions, facilitating the interpretation of the image. Thresholding and region growing are the most common algorithms used for automated segmentation, and more sophisticated segmentation techniques, such as multiresolution analysis and the Markov random field (MRF) model, are also used to enhance segmentation. However, these methods are mainly designed for a single modality, such as MRI alone. For the segmentation of brain structure, the typical approach is atlas-based registration after preprocessing and spatial normalization of individual brain scans (14). The atlas, containing labeled segmentations, is either volume-based or surface-based and exists in a specific templated imaging space (15-17). Subject images are registered to the template, and the atlas is applied to map the location of the labeled brain structures. However, atlas-based methods ignore individual specificity

because normal anatomical differences among patients can affect the performance of the method. The registration process can also be challenged when there are significant differences between the subject and template images, such as head tilt during image acquisition. Recently, some toolkits, such as FreeSurfer (18) and statistical parametric mapping (SPM) (19), have started to be used for segmentation tasks that can provide large-scale population-based segmentation results. Although these tools enhance segmentation efficiency, there are also expensive computational costs and potential failures in image registration (20-22). In addition, rigorous preprocessing steps, including skull stripping and bias correction, are required to increase the stability of these computational tools (13).

In recent years, with improvements in computing power, deep learning has been successfully applied in various fields, achieving good performance (23-29). Some rapid and effective deep learning-based parcellation approaches have been proposed (30-32) to overcome the limitations of indiscriminately applying atlas-based registration in healthy subjects. Encoder-decoder networks (33-35) are commonly used for medical image segmentation, such as segmenting retinal vascular datasets and analyzing the differences between the retinas of healthy individuals and those with AD based on the segmentation results (36). Guha Roy *et al.* proposed a QuickNAT model to segment the whole brain into 27 structures based on the U-Net architecture (37). Rashed *et al.* applied a single-encoder and multi-decoder network for segmenting the brain into 7 structures (38). Li *et al.* used a multi-view approach to segment the claustrum in T1-weighted MRI scans (39), obtaining better segmentation results than several methods. However, these methods are mainly designed for MRI with a modality. Some approaches have also introduced multimodal information, combining PET and MR images. Subramanyam Rallabandi and Seetharaman developed an Inception-ResNet wrapper model to differentiate healthy controls (HC), mild cognitive impairment (MCI), and AD, which takes the fusion of MR and PET images as input (40). Similarly, Kong *et al.* proposed an image fusion method to

fuse MR images with PET images from AD patients and used three-dimensional (3D) convolutional neural networks (CNNs) to evaluate the effectiveness of the proposed fusion approach in both dichotomous and multiclassification tasks (41). Both methods fuse PET and MR images together before feeding them into the network. Huang *et al.* proposed two different network architectures based on a 3D CNN to implement the multimodality classifier, which discussed the case of PET and MR images as parallel and independent inputs, respectively (23). However, this approach only targets the hippocampal area.

In this paper, we explore an automatic brain segmentation method based on ^{18}F -fluorodeoxyglucose (^{18}F -FDG) PET/MR dual-modality registration with an encoder-decoder architecture. MR images address the shortcomings of low-resolution PET images, while PET images supply functional features. To improve performance and reduce time costs, functional (PET) and structural (MRI) features are both employed in the encoder-decoder architecture. To assess our method, we investigate four canonical evaluation metrics and compare the values to those obtained with PET-only and MRI-only methods. The dual-mode input method is expected to achieve better results than the single-mode input approaches.

The contributions of this paper are summarized as follows:

- (I) We explored a multimodality brain segmentation method with PET and MR images as parallel inputs. Our method incorporates both functional and structural information, and the experimental results also prove the effectiveness of our approach, which has improved the segmentation accuracy compared to single-modality methods.
- (II) Our model has the ability to reduce the segmentation time. We calculated the computation time during the test process and found that it took only 20 s to segment the whole brain into 45 brain structures instead of over 6 hours with FreeSurfer toolkits, reducing the time consumption.
- (III) The proposed method can be applied to other multimodal tasks, such as lung cancer segmentation based on PET/computed tomography (CT) images. Multimodal segmentation methods can utilize more complementary information in image analysis, promoting segmentation accuracy.

The remainder of this article is arranged as follows: section “Methods” presents the details of our proposed network with the utilized datasets and evaluation metrics.

Next, we show the experimental results and an evaluation of our proposed method in section “Results”. Section “Discussion” describes a discussion of the findings and plans for future work. Finally, we report our conclusion in section “Conclusions”. We present this article in accordance with the MDAR reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-1114/rc>).

Methods

PET/MR data acquisition

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The Ethics Committee of Henan Provincial People’s Hospital & the People’s Hospital of Zhengzhou University approved this study. Because the study is a retrospective study of a sample or database established by the hospital, written informed consent form was omitted.

Our datasets contain ^{18}F -FDG PET/MR head images from 120 subjects that were acquired on an integrated 3.0 T PET/MRI scanner (uPMR 790, manufactured by United Imaging Healthcare, UIH, Shanghai, China). The patients’ heights ranged from 1.14 to 1.85 m, and the weights ranged from 17 to 115 kg. Patients were injected with an ^{18}F -FDG tracer after fasting for at least 6 hours, and the ordered subset expectation maximization (OSEM) algorithm was used to reconstruct the images. *Table 1* shows the parameters of the PET/MRI system. The sizes of the registered ^{18}F -FDG PET and MR images were $230 \times 256 \times 176$ and $300 \times 300 \times 317.8$ mm³, respectively, while the sizes of the original ^{18}F -FDG PET and MR images were $345 \times 384 \times 264$ and $192 \times 192 \times 227$ mm³, respectively.

Image preprocessing

The image preprocessing approach includes five steps, and the first two steps, reconstruction and registration, are shown in *Figure 1*. First, we reconstructed the raw MR images to $256 \times 256 \times 256$ pixel regions with a size of 1 mm. Then, we registered the raw ^{18}F -FDG PET and decapsulated MR images via Advanced Normalization Tools (ANTs). Next, we divided all images along the y-axis to obtain axial slices. After that, z score normalization was performed on the ^{18}F -FDG PET/MR images; the mean was subtracted from each pixel value, and the pixel values were divided by the standard deviation of all pixels in the slice. We used the FreeSurfer toolkit to generate masks according

Table 1 Parameters of the MRI and PET systems

| Parameter | MRI | PET |
|-------------------------------|---------------------------|-----------------------|
| Repetition time (ms) | 7.19 | – |
| Echo time (ms) | 3 | – |
| Inversion time (ms) | 750 | – |
| Width (mm), mean [SD] | 230 [345] | 300 [192] |
| Height (mm), mean [SD] | 256 [384] | 300 [192] |
| Depth (mm), mean [SD] | 176 [264] | 317.8 [227] |
| Pixel size (mm ³) | 0.666667×0.666667×0.66665 | 1.5625×1.5625×1.39999 |

MRI, magnetic resonance imaging; PET, positron emission tomography; SD, standard deviation.

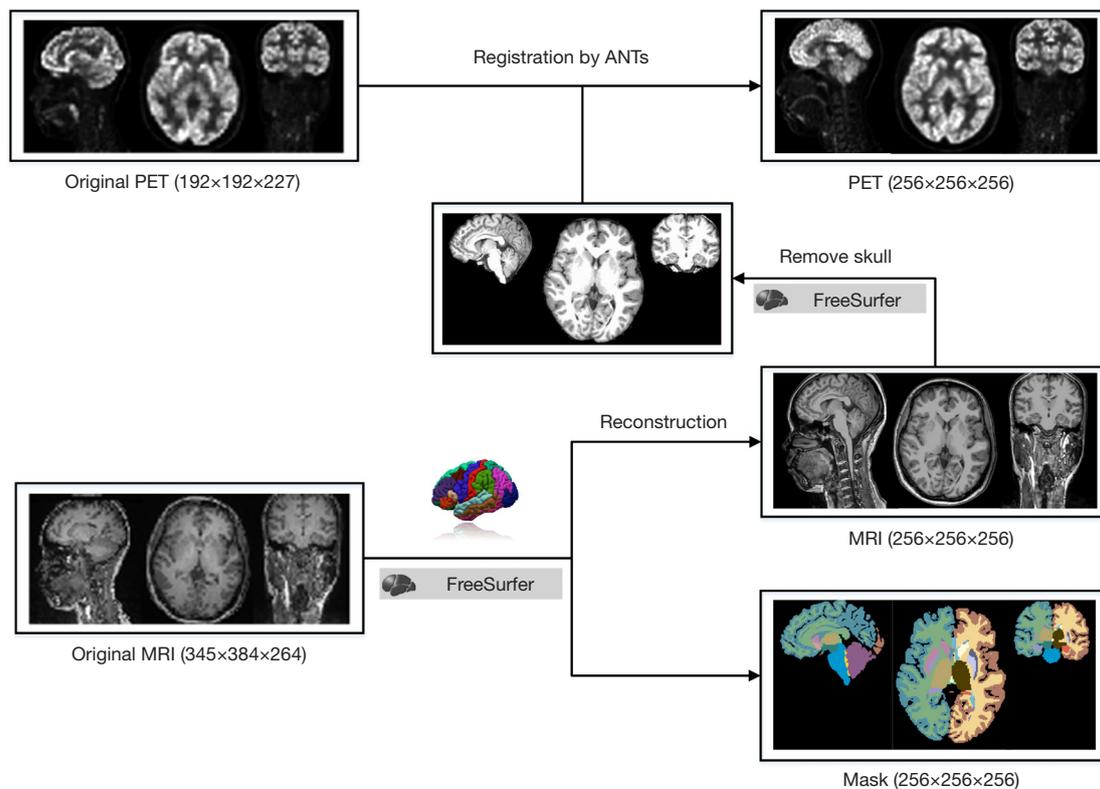


Figure 1 Reconstruction and registration by FreeSurfer and ANTs toolkits. The FreeSurfer toolkit helped to segment the raw brain MR images, generating 45 labels as the ground truth. The ANTs toolkit registered ¹⁸F-FDG PET images with MR images, and eventually, we obtained masks, ¹⁸F-FDG PET images, and MR images of the same size. ANTs, Advanced Normalization Tools; FDG, fluorodeoxyglucose; MR, magnetic resonance; MRI, magnetic resonance imaging; PET, positron emission tomography.

to atlases in standard space. The masks we obtained were applied as the ground truth during the training process. Since the ¹⁸F-FDG PET and MR images were already registered, the ground truth was the same. Finally, for

multiclass segmentation, we changed each mask slice from 256×256 to 45×256×256 with a one-hot encoder, where 45 indicated 45 different types of labels, and the labels are shown in the Table S1.

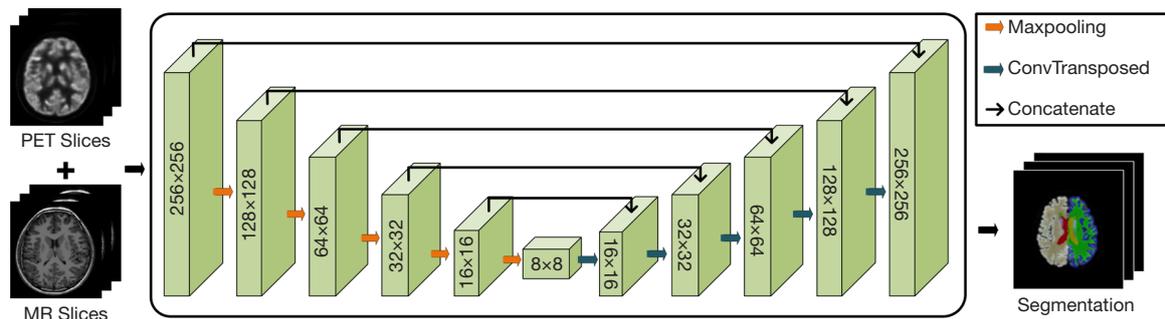


Figure 2 Illustration of our network. The network takes the ¹⁸F-FDG PET/MR slices as parallel input, and the output is the segmented mask. The U-net architecture performed multiscale feature extraction and fusion for ¹⁸F-FDG PET/MR images, facilitating accurate segmentation. FDG, fluorodeoxyglucose; MR, magnetic resonance; PET, positron emission tomography.

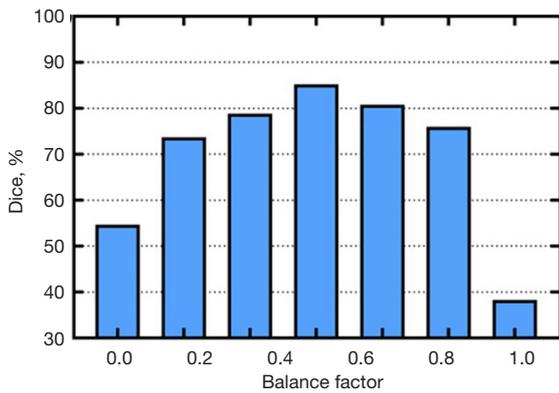


Figure 3 Dice score of the results with different balance factors.

CNN implementation

We combined the functional characteristics of PET images and the structural characteristics of MR images with CNNs to improve the segmentation results. As shown in *Figure 2*, our model is based on the U-Net architecture, which takes the axial slices of the registered ¹⁸F-FDG PET/MR images as its input during both the training and testing processes, and the output is the brain segmentation results. To extract features at different scales, we employed the two-dimensional (2D) convolutional layer with a kernel size of 3×3 and the maxpooling layer with a kernel size of 2 for downsampling. Both of the layers have a stride of 1. In addition, we used skip connections to facilitate feature fusion. In the decoding staging, we applied the transpose convolution layer with a kernel size of 2×2 and a stride of 2 to gradually recover the size of the image.

The most commonly used loss functions in image segmentation tasks are the pixelwise cross-entropy loss

and Dice coefficient loss functions. The generalized Dice loss (42) was proposed to optimize multiclass segmentation networks based on the Dice coefficient and can be formulated as follows:

$$L_{dice} = -\frac{1}{M} \sum_{j=1}^M \left(\frac{2 \sum_{i=1}^N p_{i,j} g_{i,j} + 1}{\sum_{i=1}^N p_{i,j} + \sum_{i=1}^N g_{i,j} + 1} \right) \tag{1}$$

where M represents the number of classes, N represents the number of pixels, $p_{i,j}$ represents the j^{th} class of the i^{th} pixel in the predicted slice, and $g_{i,j}$ represents the j^{th} class of the i^{th} pixel in the ground truth slice.

Moreover, the cross-entropy loss function (43,44) is the most widely used loss function in image segmentation tasks and is suitable for both binary classification and multiclass classification tasks. For multiclass segmentation tasks, the loss function can be formulated as follows:

$$L_{ce} = \sum_{i=1}^N \sum_{j=1}^M -g_{i,j} \log(p_{i,j}) \tag{2}$$

In view of the above considerations, we used a mixed loss function with both the Dice loss and cross-entropy loss. Our proposed loss function is formulated as follows:

$$Loss = \alpha L_{dice} + (1 - \alpha) L_{ce} \tag{3}$$

Where α is known as the balance factor and is the hyperparameter of the weight L_{dice} while $(1 - \alpha)$ is the hyperparameter of the weight L_{ce} .

To determine the optimal parameter value of α , we set this hyperparameter to 0, 0.2, 0.4, 0.5, 0.6, 0.8, and 1.0. The Dice coefficient was used as the evaluation metric, and the experiments were conducted under the same conditions. The quantification results are shown in *Figure 3*. We found that when the value of α is 0 or 1.0, the Dice value was

much lower, which indicates that combining multiple loss functions helps in the segmentation of brain structures. In particular, the model worked best when the value of α was equal to 0.5. Therefore, we set α to 0.5 in our experiments.

Our algorithm was implemented in the PyTorch framework with the Adam optimizer. The initial learning rate was 1×10^{-4} and decreased to 1×10^{-6} after 300 epochs. In addition, we set the batch size to 16. The training procedures of the networks were conducted on a personal computer with an NVIDIA GTX3090Ti GPU.

Evaluation metrics

Four metrics were used to evaluate different aspects of the segmentation performance in the reported experiments. The Dice similarity is an overlap metric that is commonly used to quantify segmentation accuracy, and it is formulated as follows:

$$Dice = \frac{2|V_{seg} \cap V_{gt}|}{|V_{seg}| + |V_{gt}|} = \frac{2TP}{2TP + FP + FN} \quad [4]$$

where V_{seg} denotes the pixels in the predicted binary segmentation result, while V_{gt} are the pixels in the ground-truth binary segmentation result. For a certain category, the pixels that belong to this category are positive, and those that do not belong to this category are negative. TP denotes the number of positive ground-truth pixels for which the predicted pixels are also positive. FP is the number of negative ground-truth pixels for which the predicted pixels are positive. FN is the number of positive ground-truth pixels for which the predicted pixels are negative.

The Jaccard coefficient is typically used to assess the degree of similarity between two sets (sometimes referred to as the intersection over the union or IoU). It is defined as:

$$Jaccard = \frac{|V_{seg} \cap V_{gt}|}{|V_{seg} \cup V_{gt}|} = \frac{TP}{TP + FP + FN} \quad [5]$$

Precision represents the proportion of positive pixels with correct predictions, which refers to samples with positive predicted values and positive ground-truth values, defined as follows:

$$Precision = \frac{TP}{TP + FP} \quad [6]$$

The final metric is sensitivity, which describes the proportion of identified positive cases to the total number of positive cases.

$$Sensitivity = \frac{TP}{TP + FN} \quad [7]$$

In summary, we selected these four evaluation metrics to measure the model performance from different perspectives and compared the single-modality and dual-modality approaches through the quantitative results.

Results

Overall quantitative quality

We first present the overall comparative results between the dual-modality and single-modality inputs in different views, as shown in *Figure 4*. The PET-only-based results show only rough outlines, while the MR-only-based results and the proposed method show more details. Moreover, the results of the proposed method are more consistent with the ground truth than the MR image input results, such as the yellow area in *Figure 4A*. In the outlined area, the red mask is not closed in the PET-only-based segmentation results, while the green mask from the MRI-only-based methods has a gap. In contrast, the dual-modality result is closer to the ground truth. In *Figure 4B*, the PET-only-based method can only predict the overall outline, while the detailed prediction performance is poor. The MRI-only and dual-modality prediction results are almost the same as the gold standard for the boxed region, which was placed in the lower right corner, but the dual-modality results in the upper right bifurcation are closer to the ground truth. Moreover, in *Figure 4E*, we found that the proposed model can perceive the presence of small structures that cannot be successfully segmented using the high-resolution MR images alone while combining the structural and functional information. The visual results demonstrate that compared with single-modality PET, dual-modality inputs help the method to predict more details.

Figure 5 shows the values of the four evaluation indicators for all categories, which were obtained from our method. As shown in *Figure 5*, 15 labels have statistical Dice values greater than 90%, 14 labels have values greater than 80% and less than 90%, 9 labels have values between 60% and 80%, and only the 19th (left-vessel), 35th (right-vessel), 38th (non-WM-hypointensity) and 39th (optic-Chiasm) labels cannot be predicted because the test cases contain only 22.5, 12.75, 0.04 and 168.67 pixels on average out of a total of 173,056 pixels in the whole brain. We presented four cases where the Dice value is greater than

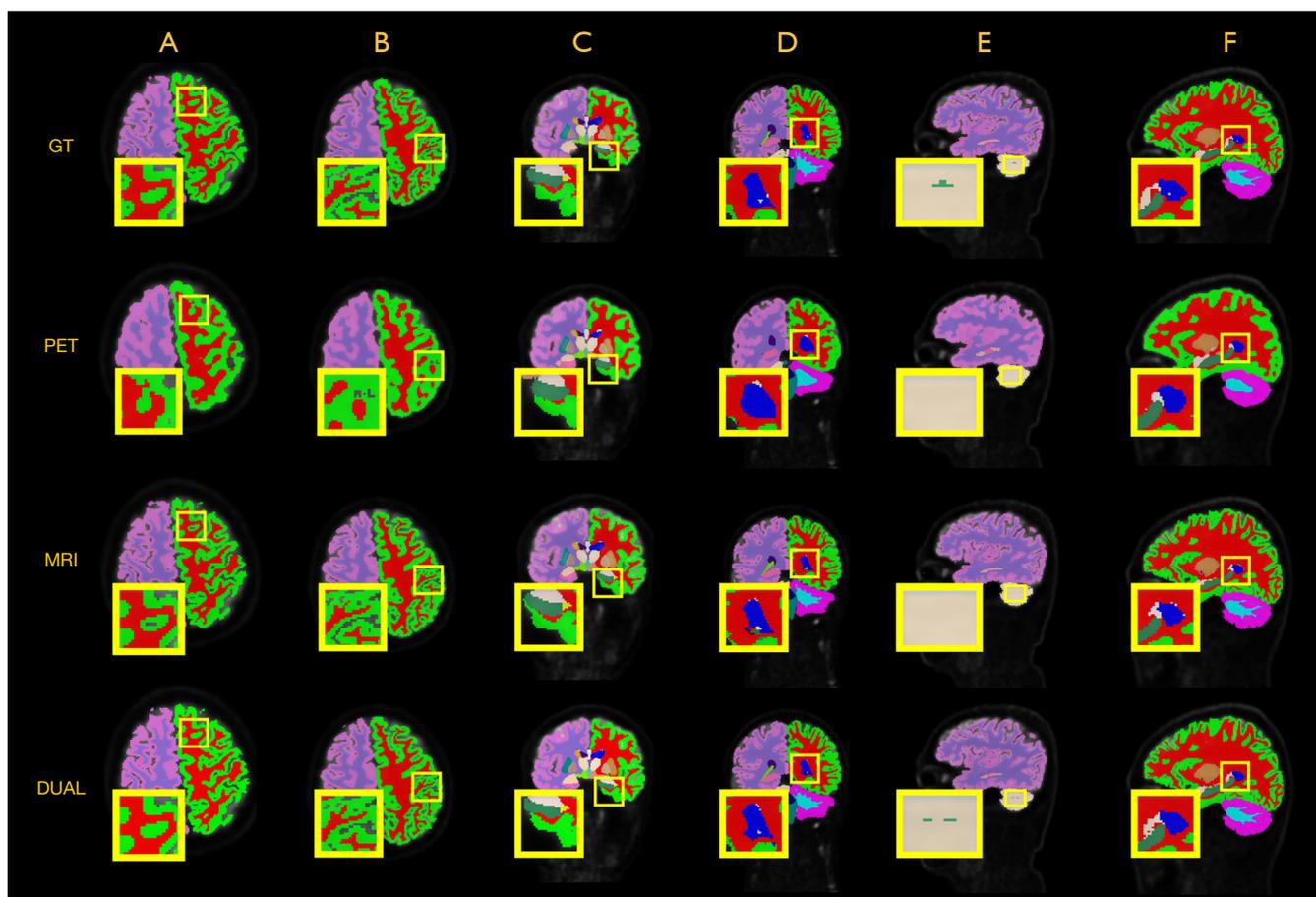


Figure 4 Segmentation results of all classes shown in axial, coronal, and sagittal views. The first row shows the ground truth, and the second row presents results from the PET-only-based segmentation, while the third row presents the results from the MRI-only-based segmentation. The last row shows the results of our method, and the yellow boxes denote the regions of interest. (A) Frontal lobe; (B) parietal lobe; (C) temporal occipital lobe; (D) lateral ventricle; (E) cerebellum; (F) corpus callosum. DUAL, dual-modality method; GT ground truth; MRI, magnetic resonance imaging; PET, positron emission tomography.

90%, between 80% and 90%, between 60% and 80%, and not predictable. In addition to the above 4 cases, the mean and standard deviation values of the Dice, Jaccard, precision and sensitivity metrics are $84.24\% \pm 1.44\%$, $74.36\% \pm 2.40\%$, $84.33\% \pm 1.56\%$ and $84.73\% \pm 1.56\%$ in the 24 tested cases, respectively.

We compare the results of the single-modality and dual-modality inputs in terms of the four evaluation indicators, and the results are shown in the box plot in *Figure 6*. The box plot has 5 components: the bottom edge, quarter digit, median, three-quarter digit and upper edge. As shown in *Figure 6*, the mean values of our proposed method are the highest for four evaluation metrics, which demonstrates that the segmentation results from the proposed dual-modality

method can obtain better quantitative results reflecting the superiority of our method over the unimodal method.

In summary, regardless of the visual results or quantitative results, the method proposed in this paper performs better than the single-channel input methods, substantially improving the segmentation effect from single PET images.

Specific brain structure quantitative quality

To better illustrate the segmentation performance of our method for complex and small brain structures, we present the segmentation results for Right-Cerebral-White-Matter, Left-Inferior-Lateral-Ventricle, Left-Lateral-

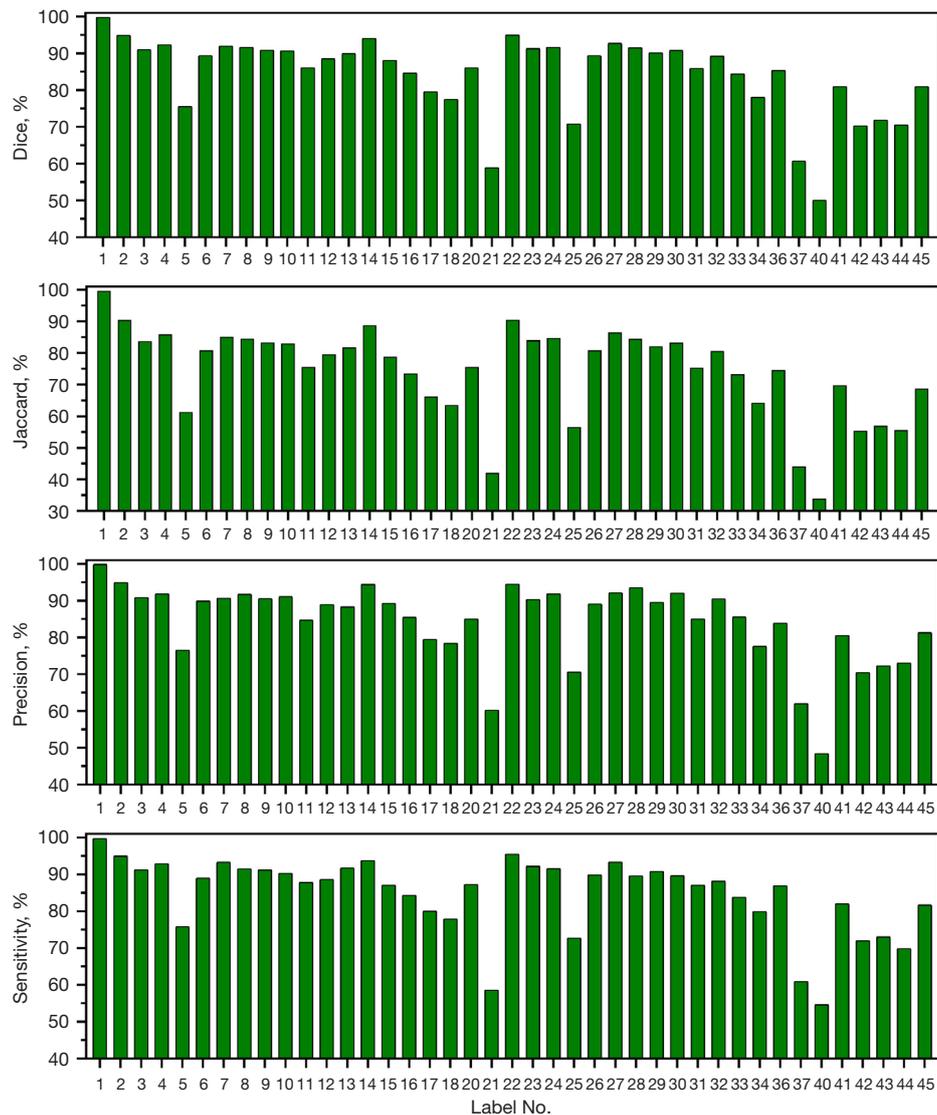


Figure 5 Quantification results for all labels. The bar charts show the specific values of the four evaluation indicators for all labels.

Ventricle and 3rd-Ventricle in *Figure 7*. The first to third rows of *Figure 7* represent three views of the segmentation results for the Right-Cerebral-White-Matter in case 1, and the yellow boxes represent the areas where our method outperforms the other methods. We find that the single-modality PET predicts only the rough outline, while the single-modality MRI and dual-modality methods are consistent with the ground truth. In particular, for the coronal view of case 1, in the yellow box, there is a very small dotted structure, namely, the occipital gyrus, in the ground truth that cannot be predicted by single-modality methods, while our method succeeds. Similarly, for the

axial and sagittal views of case 1, the results in the boxed region from the dual-modality methods are closer to the ground truth and better at predicting details. In addition, we show another example with the Left-Inferior-Lateral-Ventricle, Left-Lateral-Ventricle and 3rd-Ventricle from case 2, and the same situation occurs in these results. The visual results show that in terms of the label comparison, the results of the single-modality MRI and bimodal methods are almost the same as the true label, far exceeding the results of the single-modality PET method, and the bimodal method outperforms the single-modality MRI approach.

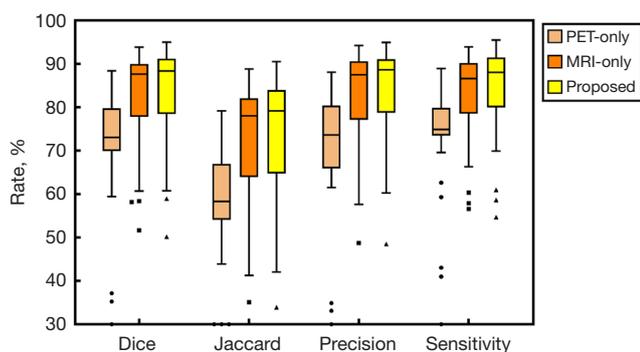


Figure 6 Box plot of the Dice, Jaccard, precision and sensitivity metrics obtained by the single-modality PET, single-modality MRI and proposed dual-modality PET/MRI methods. ‘•’, ‘■’ and ‘▲’ denote outliers. MRI, magnetic resonance imaging; PET, positron emission tomography.

Furthermore, we select five other types of labels from another case according to their Dice values to compare the results of the dual-modality and single-modality methods in the axial, coronal and sagittal views, as shown in *Figure 8*. The Dice values are the means of the labels shown in the slices. For example, in the coronal view, the segmentation results of the MRI-only and proposed approaches show more details than the PET-only segmentation results, while the segmentation results of the proposed method are closer to the ground-truth label in the blue mask and the middle pons ring area. Visually and numerically, the results of the proposed method are better than those of the MRI-only method and the PET-only method.

For a comprehensive comparison, we randomly select sixteen labels and compare the Dice values of the PET-only, MRI-only and proposed methods, as shown in *Table 2*. The best values are highlighted with bold font, and the overall result is included in the last line of the table. The proposed algorithm obtained accurate delineation, with an average Dice similarity score of 86.87%. Compared to the PET-only-based method, we improved the Dice value by 23.29% and performed slightly better than the MRI-only method. Specifically, for label Nos. 2, 8, 22 and 28, the Dice values of the MRI-only segmentation methods were below 90%, while our method exceeded 90%, greatly improving the segmentation accuracy. Additionally, for label No. 36, the PET-only-based method cannot accomplish the segmentation task with a Dice value of zero. Although the MRI-only segmentation method enables segmentation, it also has a relatively lower Dice

value, while our method obtains higher quantification results. For the 16 selected labels, the Dice values of the proposed method are the best relative to the single-modality approach, reflecting the effectiveness of fusing structural and functional information.

Clinical quantitative quality

In this subsection, the tumor background ratio was used to quantitatively evaluate the PET image segmentation, and the results are listed in *Table 3*, where the values for the ground truth and the proposed method represent the ratio of tumor background in the ground truth and dual-mode segmentation results, while the rate indicates the relative error of the tumor background ratio between the ground truth and the proposed results. The tumor background ratio is equal to the intake dose of a certain brain area compared to the intake dose of the upper pons. Here, we chose the lateral ventricle (label Nos. 3 and 23 are the left and right lateral ventricles, respectively), the thalamus (label Nos. 4 and 27 are the left and right thalamus), and the caudate nucleus (label Nos. 8 and 28 are the left and right caudate nucleus) as our brain structures of interest. We chose these regions because they have considerable clinical significance: lateral ventricle dysfunction can cause hydrocephalus, the cerebellum controls motor coordination, cerebral infarction often occurs in the thalamus, and atrophy of the caudate nucleus is associated with Huntington’s disease (HD). The experimental results show that our method has a clinical tolerance rate of no more than 5% when compared with that of the ground truth.

Discussion

This work presents an automatic segmentation algorithm based on a 2D deep neural network with two input channels. We compared our segmentation results to the results of single-input methods in terms of the Dice, Jaccard, precision and sensitivity metrics, and the results showed that the proposed method achieves more accurate and faster brain structure segmentation. The overall and individual label effects were assessed visually, and the scores of various labels in terms of four evaluation indicators were compared. In addition, the results of the proposed method have a clinical tolerance rate of less than 5% compared with the ground truth. The results show that our proposed method outperforms single-channel input methods, especially the PET single-channel input approach. Thus, our method

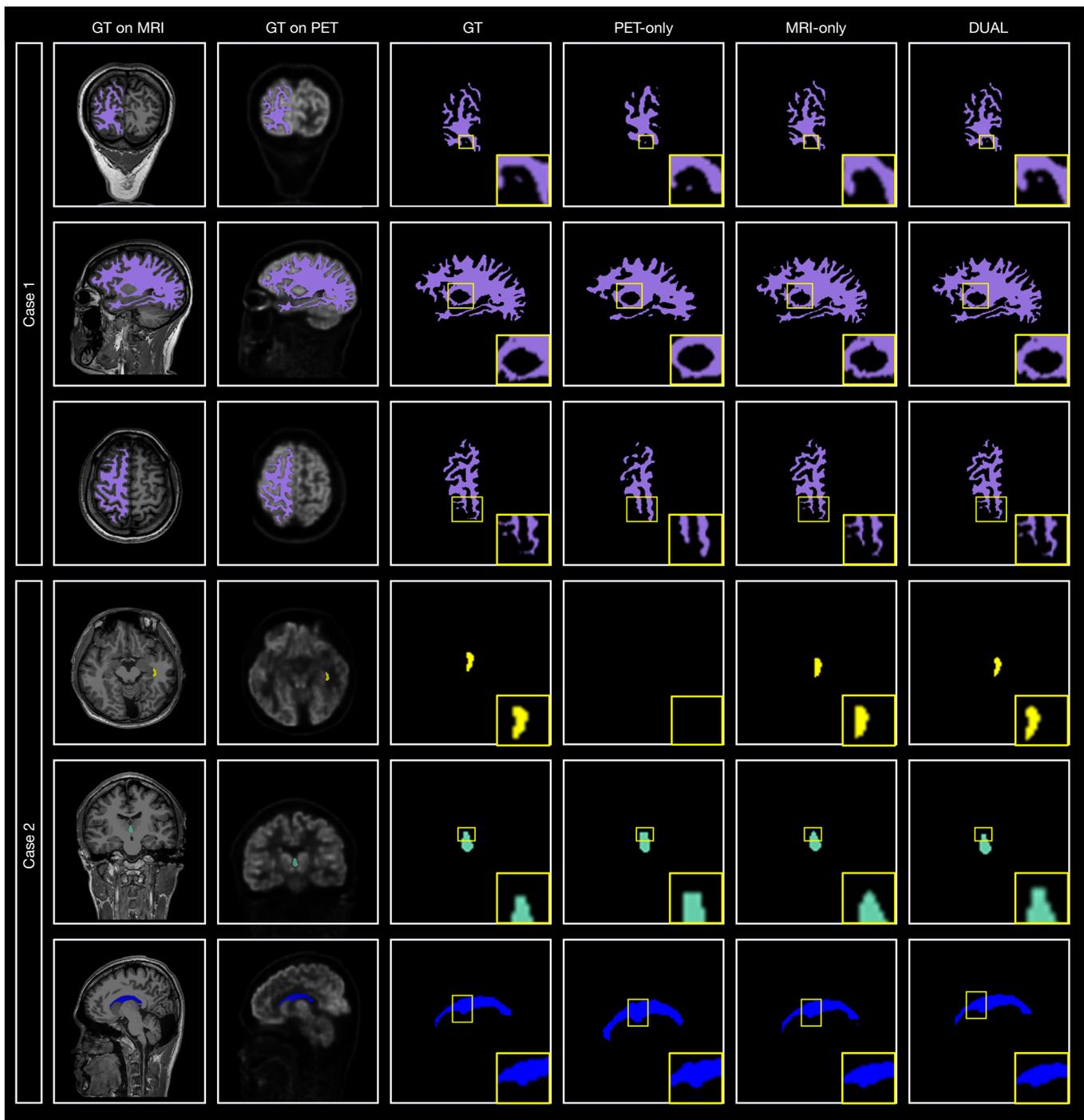


Figure 7 Segmentation results from two cases in the axial, coronal, and sagittal views. For case 1, we present the segmentation results of the Right-Cerebral-White-Matter. For case 2, we present the segmentation results of the Left-Inferior-Lateral-Ventricle, Left-Lateral-Ventricle and 3rd-Ventricle. DUAL, dual-modality method; GT, ground truth; MRI, magnetic resonance imaging; PET, positron emission tomography.

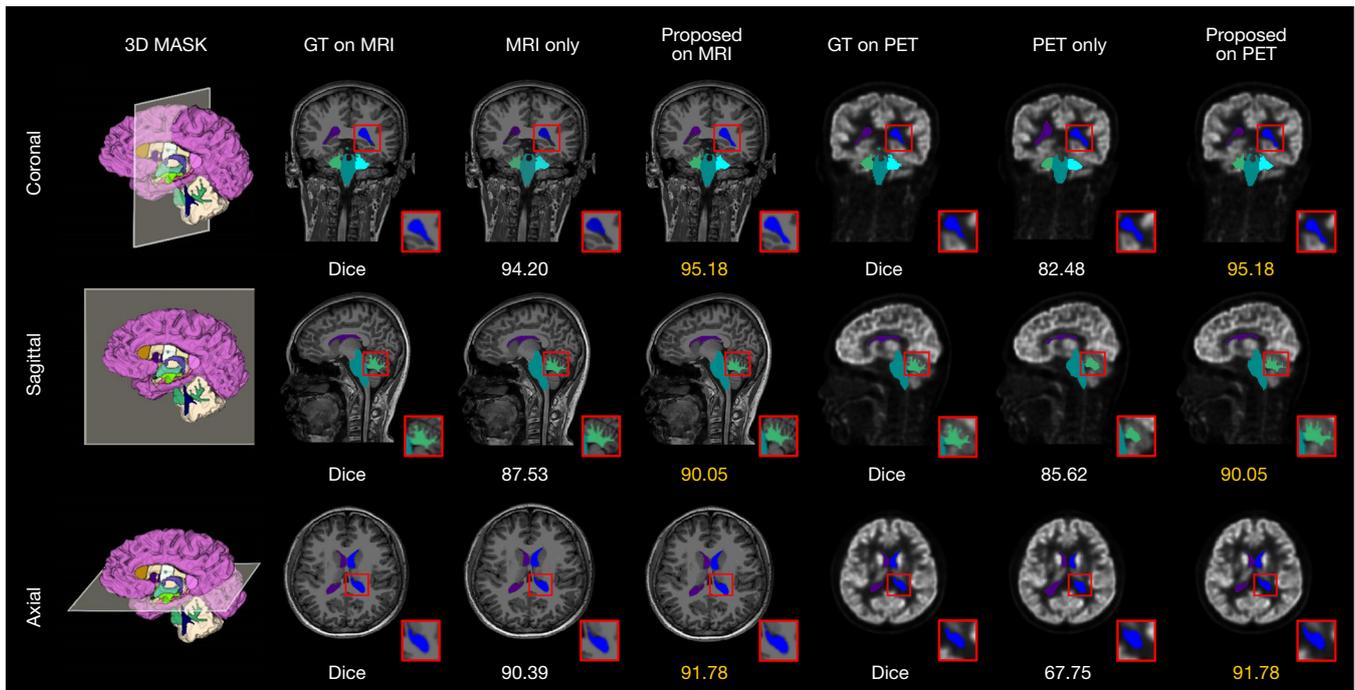


Figure 8 Segmentation results of the left lateral ventricle, left cerebellum white matter, right lateral ventricle, right cerebellum white matter and brain stem in the axial, coronal, and sagittal views. The Dice values are the mean values of the slices for the given labels, and the best results are marked in yellow. GT, ground truth; MRI, magnetic resonance imaging; PET, positron emission tomography.

Table 2 Quantitative comparison of the Dice measures obtained by the PET-only, MRI-only and proposed methods for 16 labels

| Label No. | Name | PET-only | MRI-only | Proposed method |
|-----------|----------------------------|----------|----------|-----------------|
| 1 | Left-Cerebral-White-Matter | 76.81% | 93.82% | 94.95% |
| 2 | Left-Cerebral-Cortex | 70.38% | 89.33% | 91.07% |
| 3 | Left-Lateral-Ventricle | 71.68% | 91.11% | 92.37% |
| 7 | Left-Thalamus-Proper | 80.82% | 90.80% | 91.58% |
| 8 | Left-Caudate | 80.91% | 88.61% | 90.86% |
| 10 | Left-Pallidum | 78.74% | 84.78% | 86.04% |
| 13 | Brain-Stem | 88.37% | 93.53% | 94.07% |
| 15 | Left-Amygdala | 74.44% | 82.94% | 84.65% |
| 18 | Left-Ventral DC | 71.45% | 84.99% | 86.08% |
| 22 | Right-Cerebral-Cortex | 70.93% | 89.75% | 91.26% |
| 23 | Right-Lateral-Ventricle | 69.05% | 90.32% | 91.66% |
| 27 | Right-Thalamus-Proper | 81.19% | 90.22% | 91.52% |
| 28 | Right-Caudate | 81.26% | 88.23% | 90.14% |
| 36 | Right-Choroid-Plexus | 00.00% | 58.20% | 60.77% |
| 42 | Cerebral Cortex_Central | 59.43% | 68.54% | 71.83% |
| 44 | Cerebral Cortex_Anterior | 71.92% | 78.35% | 81.02% |
| Overall | - | 70.46% | 85.22% | 86.87% |

DC, dorsal caudate; MRI, magnetic resonance imaging; PET, positron emission tomography.

Table 3 Tumor background ratio in two cases in the lateral ventricle, thalamus, and caudate nucleus

| Label No. | Name | Case 3 | | | Case 4 | | |
|-----------|---------------------------------|--------------|-----------------|-------|--------------|-----------------|-------|
| | | Ground truth | Proposed method | Rate | Ground truth | Proposed method | Rate |
| 3 | Left-Lateral-Ventricle | 0.4153 | 0.4155 | 0.04% | 0.3489 | 0.3480 | 0.24% |
| 4 | Left-Inferior-Lateral-Ventricle | 0.3220 | 0.3206 | 0.43% | 0.3325 | 0.3271 | 0.16% |
| 8 | Left-Caudate | 0.0229 | 0.0221 | 3.42% | 0.0154 | 0.0157 | 2.08% |
| 23 | Right-Lateral-Ventricle | 0.5746 | 0.5933 | 3.26% | 0.6324 | 0.6276 | 0.76% |
| 27 | Right-Thalamus-Proper | 0.3194 | 0.3165 | 0.93% | 0.3287 | 0.3389 | 3.10% |
| 28 | Right-Caudate | 0.3331 | 0.3392 | 0.18% | 0.3153 | 0.3310 | 5.00% |

provides an opportunity for personalized assessments and can be applied as a tool to analyze brain diseases with high efficiency and accuracy.

In addition, we found that the proposed method can reduce the segmentation time compared to using the FreeSurfer toolkits. During the test process, we calculated the computation time, which only required 20 seconds to segment the whole brain to 45 brain structures instead of over 6 hours with the FreeSurfer toolkit. Some studies (45–47) also suggest that the time to segment brain images is longer with the FreeSurfer toolkit. For instance, Rebsamen *et al.* spent an average of 9.3 hours for 454 MR images using FreeSurfer toolkits, and when they enhanced the contrast of MR images, the median processing runtime increased to 15.6 hours (48). Existing references indicate that traditional whole-brain segmentation methods still apply standard brain templates (7). The segmentation accuracy of these methods is limited, and tools such as FreeSurfer take a considerable amount of time to perform segmentation. In this article, we used deep learning methods and dual-modal input features to improve the segmentation performance and reduce time costs.

Class unbalancing is also worthy of attention. When we segmented the brain, we found that the 19th (left-vessel), 35th (right-vessel), 38th (non-WM-hypointensity) and 39th (optic-chiasm) labels could not be segmented because they occupied only 22.5, 12.75, 0.04 and 168.67 pixels on average out of a total of 173,056 pixels in the whole brain. Different shapes of the structures lead to the problem of class imbalance, which also poses a challenge to the segmentation task (49). In future work, we will consider adding weights to these structures that occupy fewer pixels to facilitate the segmentation of small structures.

The segmentation of some tiny structures also deserves attention. Our method cannot segment the 19th, 35th, 38th

and 39th labels because the proportion of these labels in the brain is less than 1/1,000 on average, which indicates the difficulty of segmentation for some tiny structures by our method. In future work, we will develop improvements to our method to facilitate the precise segmentation of brain images. In addition, considering that manual labeling is tedious and labor intensive (1,50), we used the brain mask generated by the FreeSurfer toolkit as ground truth. To increase the credibility of our method, we will also collect more patient data to validate our method in future work.

Conclusions

In this study, we explore a deep learning-based whole-brain automatic segmentation method based on PET/MR registration. The purpose of this work is to improve the segmentation accuracy and to quickly obtain segmentation results to assist doctors in subsequent patient diagnoses and treatment. Our experiments prove that the bimodal PET/MR method has a better segmentation effect than either mode individually. We found that (I) after registration, PET and MR images contain complementary information, with MRI information considerably assisting PET images; (II) the dual-modality segmentation results are better than the single-modality segmentation results. Although our method performs better for brain segmentation, it is still difficult to segment some tiny structures. In future work, we will improve our method and apply our method to other multimodal segmentation tasks, such as PET/CT.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (Nos. 32022042 and 62101540), the Shenzhen Excellent Technological

Innovation Talent Training Project of China (No. RCJC20200714114436080), the Shenzhen Science and Technology Program (Nos. JCYJ20220818101804009 and RCBS20210706092218043), the China Postdoctoral Science Foundation (No. 2022M713290), Shenzhen Municipal Scheme for Basic Research of China (No. JCYJ20210324100208022), and the Natural Science Foundation of Guangdong Province-Outstanding Youth Project (No. 2023B1515020002).

Footnote

Reporting Checklist: The authors have completed the MDAR reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-22-1114/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-1114/coif>). DL serves as an unpaid editorial board member of *Quantitative Imaging in Medicine and Surgery*. JY, QH and ZW are employees of the Shanghai United Imaging Healthcare Group. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The Ethics Committee of Henan Provincial People's Hospital & the People's Hospital of Zhengzhou University approved this study. Because the study is a retrospective study of a sample or database established by the hospital, written informed consent form was omitted.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Giorgio J, Jagust WJ, Baker S, Landau SM, Tino P, Kourtzi Z; . A robust and interpretable machine learning approach using multimodal biological data to predict future pathological tau accumulation. *Nat Commun* 2022;13:1887.
2. Yuan J, Ran X, Liu K, Yao C, Yao Y, Wu H, Liu Q. Machine learning applications on neuroimaging for diagnosis and prognosis of epilepsy: A review. *J Neurosci Methods* 2022;368:109441.
3. Zhang J, He X, Qing L, Gao F, Wang B. BPGAN: Brain PET synthesis from MRI using generative adversarial network for multi-modal Alzheimer's disease diagnosis. *Comput Methods Programs Biomed* 2022;217:106676.
4. Wang H, Wu Y, Huang Z, Li Z, Zhang N, Fu F, Meng N, Wang H, Zhou Y, Yang Y, Liu X, Liang D, Zheng H, Mok GSP, Wang M, Hu Z. Deep learning-based dynamic PET parametric K(i) image generation from lung static PET. *Eur Radiol* 2023;33:2676-85.
5. Huang Z, Liu Z, He P, Ren Y, Li S, Lei Y, Luo D, Liang D, Shao D, Hu Z, Zhang N. Segmentation-guided Denoising Network for Low-dose CT Imaging. *Comput Methods Programs Biomed* 2022;227:107199.
6. Bar-Sever Z, Biassoni L, Shulkin B, Kong G, Hofman MS, Lopci E, Manea I, Kozirowski J, Castellani R, Boubaker A, Lambert B, Pfluger T, Nadel H, Sharp S, Giammarile F. Guidelines on nuclear medicine imaging in neuroblastoma. *Eur J Nucl Med Mol Imaging* 2018;45:2009-24.
7. Chen Z, Qiu T, Tian Y, Feng H, Zhang Y, Wang H. Automated brain structures segmentation from PET/CT images based on landmark-constrained dual-modality atlas registration. *Phys Med Biol* 2021.
8. Täieb D, Hicks RJ, Hindié E, Guillet BA, Avram A, Ghedini P, Timmers HJ, Scott AT, Elojeimy S, Rubello D, Virgolini IJ, Fanti S, Balogova S, Pandit-Taskar N, Pacak K. European Association of Nuclear Medicine Practice Guideline/Society of Nuclear Medicine and Molecular Imaging Procedure Standard 2019 for radionuclide imaging of pheochromocytoma and paraganglioma. *Eur J Nucl Med Mol Imaging* 2019;46:2112-37.
9. Zhu L, He Q, Huang Y, Zhang Z, Zeng J, Lu L, Kong W, Zhou F. DualMMP-GAN: Dual-scale multi-modality perceptual generative adversarial network for medical image segmentation. *Comput Biol Med* 2022;144:105387.
10. Sun H, Jiang Y, Yuan J, Wang H, Liang D, Fan W, Hu Z, Zhang N. High-quality PET image synthesis from ultra-

- low-dose PET/MRI using bi-task deep learning. *Quant Imaging Med Surg* 2022;12:5326-42.
11. Mawlawi O, Townsend DW. Multimodality imaging: an update on PET/CT technology. *Eur J Nucl Med Mol Imaging* 2009;36 Suppl 1:S15-29.
 12. Zheng P, Zhu X, Guo W. Brain tumour segmentation based on an improved U-Net. *BMC Med Imaging* 2022;22:199.
 13. Bagci U, Udupa JK, Mendhiratta N, Foster B, Xu Z, Yao J, Chen X, Mollura DJ. Joint segmentation of anatomical and functional images: applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images. *Med Image Anal* 2013;17:929-45.
 14. Liu Y, Wei Y, Wang C. Subcortical Brain Segmentation Based on Atlas Registration and Linearized Kernel Sparse Representative Classifier. *IEEE Access* 2019;7:31547-57.
 15. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 2006;31:968-80.
 16. Frazier JA, Chiu S, Breeze JL, Makris N, Lange N, Kennedy DN, Herbert MR, Bent EK, Koneru VK, Dieterich ME, Hodge SM, Rauch SL, Grant PE, Cohen BM, Seidman LJ, Caviness VS, Biederman J. Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *Am J Psychiatry* 2005;162:1256-65.
 17. Makris N, Goldstein JM, Kennedy D, Hodge SM, Caviness VS, Faraone SV, Tsuang MT, Seidman LJ. Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophr Res* 2006;83:155-71.
 18. Greve DN, Salat DH, Bowen SL, Izquierdo-Garcia D, Schultz AP, Catana C, Becker JA, Svarer C, Knudsen GM, Sperling RA, Johnson KA. Different partial volume correction methods lead to different conclusions: An (18)F-FDG-PET study of aging. *Neuroimage* 2016;132:334-43.
 19. Friston K, Ashburner J, Kiebel S, Nichols TE, Penny W. editors. *Statistical Parametric Mapping*. Amsterdam: Elsevier Ltd., 2007.
 20. Baniyasi M, Petersen MV, Gonçalves J, Horn A, Vlasov V, Hertel F, Husch A. DBSegment: Fast and robust segmentation of deep brain structures considering domain generalization. *Hum Brain Mapp* 2023;44:762-78.
 21. Huang Z, Wu Y, Fu F, Meng N, Gu F, Wu Q, Zhou Y, Yang Y, Liu X, Zheng H, Liang D, Wang M, Hu Z. Parametric image generation with the uEXPLORER total-body PET/CT system through deep learning. *Eur J Nucl Med Mol Imaging* 2022;49:2482-92.
 22. Harkey T, Baker D, Hagen J, Scott H, Palys V. Practical methods for segmentation and calculation of brain volume and intracranial volume: a guide and comparison. *Quant Imaging Med Surg* 2022;12:3748-61.
 23. Huang Y, Xu J, Zhou Y, Tong T, Zhuang X; . Diagnosis of Alzheimer's Disease via Multi-Modality 3D Convolutional Neural Network. *Front Neurosci* 2019;13:509.
 24. Huang Z, Liu X, Wang R, Chen J, Lu P, Zhang Q, Jiang C, Yang Y, Liu X, Zheng H, Liang D, Hu Z. Considering anatomical prior information for low-dose CT image enhancement using attribute-augmented Wasserstein generative adversarial networks. *Neurocomputing* 2021;428:104-15.
 25. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 2017;36:61-78.
 26. Prados F, Ashburner J, Blaiotta C, Brosch T, Carballido-Gamio J, Cardoso MJ, et al. Spinal cord grey matter segmentation challenge. *Neuroimage* 2017;152:312-29.
 27. Thai A, Bui V, Reyes L, Chang LC. Using Deep Convolutional Neural Network for Mouse Brain Segmentation in DT-MRI. 2019 IEEE International Conference on Big Data (Big Data); Los Angeles, CA, USA. IEEE, 2019:6229-31.
 28. Wachinger C, Reuter M, Klein T. DeepNAT: Deep convolutional neural network for segmenting neuroanatomy. *Neuroimage* 2018;170:434-45.
 29. Chen W, Huang Z, Liu D, Lu P, Chen J. Towards Information Diversity Through Separable Cascade Modules for Image Super Resolution. 2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS); Shenyang, China. IEEE, 2021:125-31.
 30. Doyen S, Nicholas P, Poologaindran A, Crawford L, Young IM, Romero-Garcia R, Sughrue ME. Connectivity-based parcellation of normal and anatomically distorted human cerebral cortex. *Hum Brain Mapp* 2022;43:1358-69.
 31. Huang Z, Chen Z, Chen J, Lu P, Quan G, Du Y, Li C, Gu Z, Yang Y, Liu X, Zheng H, Liang D, Hu Z. DaNet: dose-

- aware network embedded with dose-level estimation for low-dose CT imaging. *Phys Med Biol* 2021;66:015005.
32. Huang Z, Liu X, Wang R, Chen Z, Yang Y, Liu X, Zheng H, Liang D, Hu Z. Learning a Deep CNN Denoising Approach Using Anatomical Prior Information Implemented With Attention Mechanism for Low-Dose CT Imaging on Clinical Patient Data From Multiple Anatomical Sites. *IEEE J Biomed Health Inform* 2021;25:3416-27.
 33. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A. editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. 2015:234-41.
 34. Huang Z, Chen Z, Quan G, Du Y, Yang Y, Liu X, Zheng H, Liang D, Hu Z. Deep Cascade Residual Networks (DCRN): Optimizing an Encoder–Decoder Convolutional Neural Network for Low-Dose CT Imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences* 2022;6:829-40.
 35. Huang Z, Liu D, Chen W, Lu P, Chen J. Adversarial Learning for Image Super Resolution Using Auxiliary Texture Feature Attributes. 2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS); Shenyang, China. *IEEE*, 2021:132-7.
 36. Ma Y, Hao H, Xie J, Fu H, Zhang J, Yang J, Wang Z, Liu J, Zheng Y, Zhao Y. ROSE: A Retinal OCT-Angiography Vessel Segmentation Dataset and New Model. *IEEE Trans Med Imaging* 2021;40:928-39.
 37. Guha Roy A, Conjeti S, Navab N, Wachinger C; . QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy. *Neuroimage* 2019;186:713-27.
 38. Rashed EA, Gomez-Tames J, Hirata A. End-to-end semantic segmentation of personalized deep brain structures for non-invasive brain stimulation. *Neural Netw* 2020;125:233-44.
 39. Li H, Menegaux A, Schmitz-Koep B, Neubauer A, Bäuerlein FJB, Shit S, Sorg C, Menze B, Hedderich D. Automated claustrum segmentation in human brain MRI using deep learning. *Hum Brain Mapp* 2021;42:5862-72.
 40. Subramanyam Rallabandi VP, Seetharaman K. Deep learning-based classification of healthy aging controls, mild cognitive impairment and Alzheimer's disease using fusion of MRI-PET imaging. *Biomedical Signal Processing and Control* 2023;80:104312.
 41. Kong Z, Zhang M, Zhu W, Yi Y, Wang T, Zhang B. Multi-modal data Alzheimer's disease detection based on 3D convolution. *Biomedical Signal Processing and Control* 2022;75:103565.
 42. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2017)* 2017;2017:240-8.
 43. Huang Z, Liu X, Wang R, Zhang M, Zeng X, Liu J, Yang Y, Liu X, Zheng H, Liang D, Hu Z. FaNet: fast assessment network for the novel coronavirus (COVID-19) pneumonia based on 3D CT imaging and clinical symptoms. *Appl Intell (Dordr)* 2021;51:2838-49.
 44. Ma YD, Liu Q, Qian ZB. Automated image segmentation using improved PCNN model based on cross-entropy. *Proceedings of the 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004; Hong Kong, China. IEEE*, 2004:743-6.
 45. Cardinale F, Chinnici G, Bramerio M, Mai R, Sartori I, Cossu M, Lo Russo G, Castana L, Colombo N, Caborni C, De Momi E, Ferrigno G. Validation of FreeSurfer-estimated brain cortical thickness: comparison with histologic measurements. *Neuroinformatics* 2014;12:535-42.
 46. Sämann PG, Iglesias JE, Gutman B, Grotegerd D, Leenings R, Flint C, Dannowski U, Clarke-Rubright EK, Morey RA, van Erp TGM, Whelan CD, Han LKM, van Velzen LS, Cao B, Augustinack JC, Thompson PM, Jahanshad N, Schmaal L. FreeSurfer-based segmentation of hippocampal subfields: A review of methods and applications, with a novel quality control procedure for ENIGMA studies and other collaborative efforts. *Hum Brain Mapp* 2022;43:207-33.
 47. Lidauer K, Pulli EP, Copeland A, Silver E, Kumpulainen V, Hashempour N, Merisaari H, Saunavaara J, Parkkola R, Lähdesmäki T, Saukko E, Nolvi S, Kataja EL, Karlsson L, Karlsson H, Tuulari JJ. Subcortical and hippocampal brain segmentation in 5-year-old children: Validation of FSL-FIRST and FreeSurfer against manual segmentation. *Eur J Neurosci* 2022;56:4619-41.
 48. Rebsamen M, McKinley R, Radojewski P, Pistor M, Friedli C, Hoepner R, Salmen A, Chan A, Reyes M, Wagner F, Wiest R, Rummel C. Reliable brain morphometry from contrast-enhanced T1w-MRI in patients with multiple sclerosis. *Hum Brain Mapp* 2023;44:970-9.

49. Pang ZF, Geng M, Zhang L, Zhou Y, Zeng T, Zheng L, Zhang N, Liang D, Zheng H, Dai Y, Huang Z, Hu Z. Adaptive weighted curvature-based active contour for ultrasonic and 3T/5T MR image segmentation. *Signal Processing*. 2023;205:108881.
50. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 2002;33:341-55.

Cite this article as: Huang Z, Liu H, Wu Y, Li W, Liu J, Wu R, Yuan J, He Q, Wang Z, Zhang K, Liang D, Hu Z, Wang M, Zhang N. Automatic brain structure segmentation for ¹⁸F-fluorodeoxyglucose positron emission tomography/magnetic resonance images via deep learning. *Quant Imaging Med Surg* 2023;13(7):4447-4462. doi: 10.21037/qims-22-1114