Improved automatic segmentation of brain metastasis gross tumor volume in computed tomography images for radiotherapy: a position attention module for U-Net architecture

Yiren Wang[1,2]^, Yiheng Hu[3], Shouying Chen[1,2], Hairui Deng[1,2], Zhongjian Wen[1,2], Yongcheng He[4], Huaiwen Zhang[5], Ping Zhou[2,6,7], Haowen Pang[8]

[1]School of Nursing, Southwest Medical University, Luzhou, China; [2]Wound Healing Basic Research and Clinical Application Key Laboratory of Luzhou, School of Nursing, Southwest Medical University, Luzhou, China; [3]Department of Medical Imaging, Southwest Medical University, Luzhou, China; [4]Department of Pharmacy, Sichuan Agricultural University, Chengdu, China; [5]Department of Radiotherapy, Jiangxi Cancer Hospital, The Second Affiliated Hospital of Nanchang Medical College, Jiangxi Clinical Research Center for Cancer, Nanchang, China; [6]Department of Nursing, The Affiliated Hospital of Southwest Medical University, Luzhou, China; [7]Department of Radiology, The Affiliated Hospital of Southwest Medical University, Luzhou, China; [8]Department of Oncology, The Affiliated Hospital of Southwest Medical University, Luzhou, China

Contributions: (I) Conception and design: P Zhou, H Pang, Y Wang; (II) Administrative support: P Zhou, H Pang; (III) Provision of study materials or patients: H Zhang, Y He, Y Wang, Y Hu, S Chen; (IV) Collection and assembly of data: Y Wang, Y Hu, S Chen, H Deng, Z Wen, Y He; (V) Data analysis and interpretation: Y Wang, Y Hu, S Chen, H Deng, Z Wen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Ping Zhou, MD, PhD. Department of Radiology, The Affiliated Hospital of Southwest Medical University, No. 25 Taiping Street, Jiangyang District, Luzhou 646000, China; Wound Healing Basic Research and Clinical Application Key Laboratory of Luzhou, School of Nursing, Southwest Medical University, Luzhou, China; Department of Nursing, The Affiliated Hospital of Southwest Medical University, Luzhou, China. Email: zhouping11@swmu.edu.cn; Haowen Pang, MD, PhD. Department of Oncology, The Affiliated Hospital of Southwest Medical University, No. 25 Taiping Street, Jiangyang District, Luzhou 646000, China. Email: haowenpang@foxmail.com.

Background: Brain metastases present significant challenges in radiotherapy due to the need for precise tumor delineation. Traditional methods often lack the efficiency and accuracy required for optimal treatment planning. This paper proposes an improved U-Net model that uses a position attention module (PAM) for automated segmentation of gross tumor volumes (GTVs) in computed tomography (CT) simulation images of patients with brain metastases to improve the efficiency and accuracy of radiotherapy planning and segmentation.

Methods: We retrospectively collected CT simulation imaging datasets of patients with brain metastases from two centers, which were designated as the training and external validation datasets. The U-Net architecture was enhanced by incorporating a PAM into the transition layer, which improved the automated segmentation capability of the U-Net model. With cross-entropy loss employed as the loss function, the samples from the training dataset underwent training. The model's segmentation performance on the external validation dataset was assessed using metrics including the Dice similarity coefficient (DSC), intersection over union (IoU), accuracy, sensitivity, specificity, Matthews correlation coefficient (MCC), and Hausdorff distance (HD).

Results: The proposed automated segmentation model demonstrated promising performance on the external validation dataset, achieving a DSC of 0.753±0.172. In terms of evaluation metrics (including the DSC, IoU, accuracy, sensitivity, MCC, and HD), the model outperformed the standard U-Net, which had

---

^ ORCID: 0000-0001-6757-5923.

a DSC of 0.691±0.142. The proposed model produced segmentation results that were closer to the ground truth and could reveal more detailed features of brain metastases.

Conclusions: The PAM-improved U-Net model offers considerable advantages in the automated segmentation of the GTV in CT simulation images for patients with brain metastases. Its superior performance in comparison with the standard U-Net model supports its potential for streamlining and improving the accuracy of radiotherapy. With its ability to produce segmentation results consistent with the ground truth, the proposed model holds promise for clinical adoption and provides a reference for radiation oncologists to make more informed GTV segmentation decisions.

Keywords: Deep learning; position attention block; artificial intelligence; brain metastases; automatic segmentation

Introduction

Brain metastases are formed by malignant tumors that originate from other parts of the body and metastasize to the brain (1). Approximately 20–40% of patients with lung and breast cancer develop brain metastases, which often lead to severe neurological impairment (2). Patients with brain metastases have a poor prognosis, making early detection and treatment particularly critical (3).

Radiotherapy is a key modality for the treatment of brain metastases (4), and accurate segmentation of the tumor volume is essential to ensuring precise and effective treatment (5). Computed tomography (CT) has become a widely used medical imaging technique for radiation therapy simulations (6). It provides clinicians and radiation physicists clear, high-resolution images that assist in determining the location, size, and shape of tumors (7). However, traditional image segmentation techniques often require that radiologists o spend a considerable amount of time and effort manually delineating the tumor boundaries (8). This approach is inefficient and can also result in inconsistencies owing to variability between operators or in small brain metastatic regions being missed (9). Consequently, the automatic and accurate segmentation of the gross tumor volume (GTV) in CT simulation images is crucial for subsequent radiation therapy.

Deep learning methods, particularly the use of convolutional neural networks (CNNs), have shown immense potential in medical image analysis (10). Previous studies have employed CNNs to automatically segment brain metastases using magnetic resonance imaging (MRI), achieving promising segmentation results (11,12). Almost all automatic segmentation studies on the GTV of brain metastases have been conducted using MRI. However, radiotherapy simulation positioning and planning are mostly determined based on simulated CT scans specific to radiotherapy (13,14). Even when MRI scans are used for fusion delineation, a certain degree of deviation exists. Consequently, the direct automatic segmentation of the GTV on simulated CT scans specific to radiotherapy has greater clinical practicality (15).

The U-Net model is a classic network designed for image segmentation tasks. Its impressive performance with small-scale datasets has garnered considerable attention in the field of medical image processing (16). The flexibility and modular design of the U-Net architecture allows for its seamless integration with other network architectures or modules, offering a pathway for further enhancement and customization. A previous study used the residual module to replace the original convolution module of U-Net in order to speed up the convergence of the model, which demonstrated faster model convergence efficiency in the segmentation of liver CT images (17). In recent years, in order to retain the features of small targets in the deep network, previous studies have also adopted the squeezing and excitation module for various image-processing tasks, which has improved the segmentation effect compared with that of the state-of-the art model (18,19). In addition, researchers have further extended the capabilities of the original U-Net by combining U-Net with Transformer in order to fully utilize the low-level features, which in turn can enhance the global features and reduce the semantic gap between the encoding and decoding stages (20). However, these methods suffer from the problems of excessive

parameter amount, with the large number of samples required for training the model hindering convergence, which limits the scope of potential applications in radiotherapy image segmentation.

In this study, we introduced a position attention module (PAM) into the transition layer of the U-Net structure. This addition allows the model to place attentional weights on features across channels, thereby enabling it to focus on spatial contextual information. The PAM learns to extract contextual information from spatial dimensions throughout the training process (21). By adding a minimal number of parameters, the computational precision and overall model performance in brain metastasis CT images can be improved. We present this article in accordance with the TRIPOD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-23-1627/rc).

Methods

Dataset preparation

This retrospective study collected the head CT images of 123 patients with brain metastases who underwent radiation therapy at the Second Affiliated Hospital of Nanchang University between January 2017 and January 2021. Additionally, head CT images of 45 patients with brain metastases treated with radiation therapy at The Affiliated Hospital of Southwest Medical University between January 2020 and January 2021 were included. This study was approved by the ethics review committees of The Affiliated Hospital of Southwest Medical University (No. KY2023041) and Jiangxi Cancer Hospital (No. 2023KY082). Owing to the retrospective nature of the study, the requirement for informed consent was waived. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The inclusion criteria were as follows: (I) age over 18 years, (II) no other brain lesions apart from brain metastases, and (III) complete CT imaging records. The exclusion criteria were as follows: (I) images with damage or artifacts, including metal shadows; (II) missing patient data, and (III) the presence of other brain lesions.

Scanning parameters

CT images from The Second Affiliated Hospital of Nanchang University were acquired using radiation therapy-positioned CT scans. The scans were performed using the SOMATOM Definition AS 20-slice CT simulator system (Siemens Healthineers, Erlangen, Germany) under the following scanning parameters: tube voltage, 120 kVp; tube current, 540 mAs; and a scanning range from the top of the skull to the third cervical vertebra. The image size was 512×512 pixels, and the scanning slice thickness was set at 3 mm with a field of view (FOV) of 250–400 mm.

CT images from the cohort of The Affiliated Hospital of Southwest Medical University were acquired from radiotherapy localization CT scans performed before radiotherapy. Radiotherapy localization CT was performed using a LightSpeed RT 4 scanner (GE HealthCare, Chicago, IL, USA). The scanning parameters used were as follows: tube voltage, 120 kVp; FOV, 250–400 mm; image size, 512×512 pixels; and slice thickness, 3 mm.

Image preprocessing and region of interest segmentation

All images were subjected to histogram equalization and median filtering to reduce the noise. The images were augmented by flipping and rotation for data enhancement. Before analysis, all images were resampled to a voxel size of $1×1×1$ mm^3 and saved in *.NII format. Before being saved, the images were deidentified to protect patient privacy. Segmentation of regions of interest (ROIs) in all images was jointly performed by two radiation therapists and radiation physicists, each with over five years of experience. In cases of disagreement, the decision was made by another two radiation therapists and physicists, each with more than 10 years of experience. The regions segmented by radiation therapists and radiation physicists served as the ground truth for this study.

PAM-improved U-Net architecture

In this study, we used the U-Net model as the foundational model framework. By incorporating the PAM into the U-Net transition layer, we aimed to enhance the performance of the standard U-Net model. The architecture of the U-Net model used in this study is shown in *Figure 1*.

Downsampling section

The images first underwent two convolution and activation operations to convert the original 3-channel image into 64 channels. The initial three channels are three consecutive axial sections of CT imaging. This approach imparts extended spatial context understanding on the model, thus facilitating enhanced recognition of anatomical and pathological patterns that persist through adjacent cross
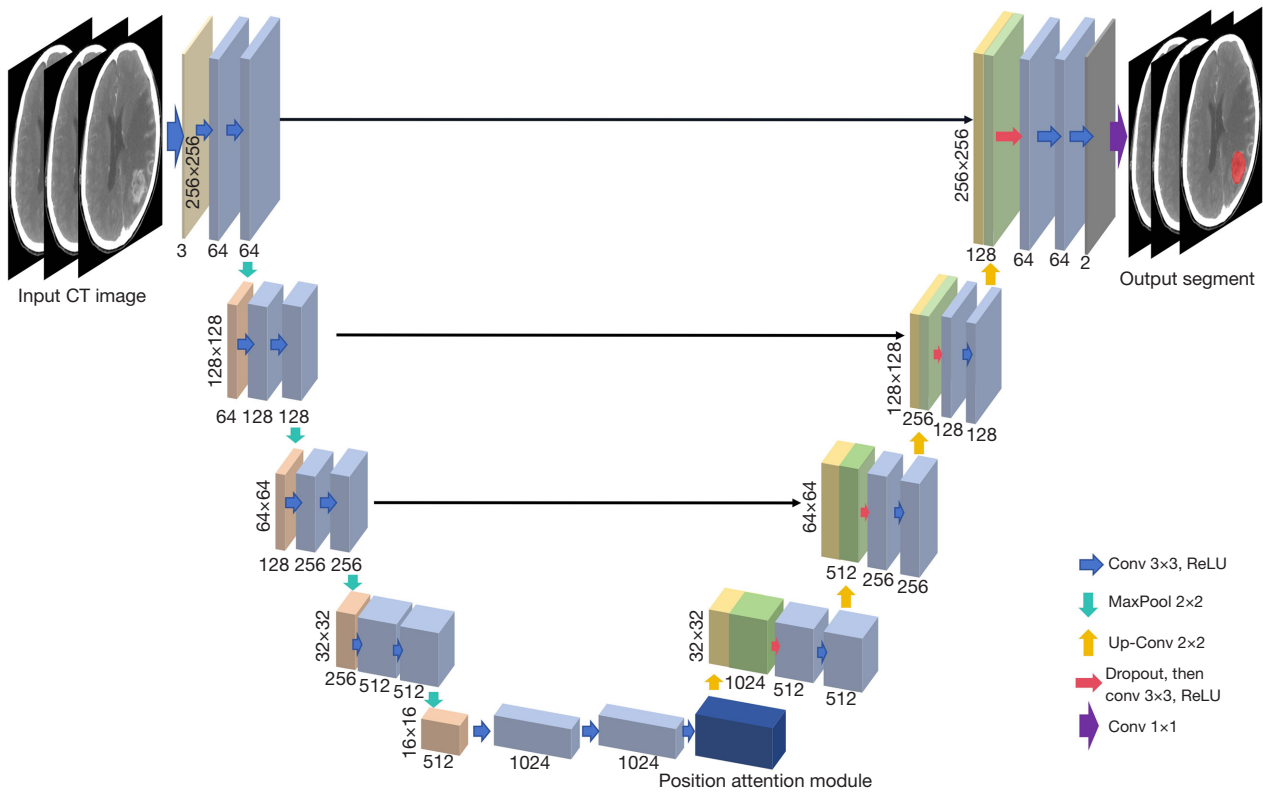
Figure 1 Architecture of the PAM-improved U-Net model. CT, computed tomography; Conv, convolution; ReLU, rectified linear unit; PAM, position attention module.

sections. Subsequently, a downsampling pooling operation was performed, reducing the image size from 512×512 to 256×256 pixels while increasing the feature map channels to 128. This downsampling process was executed four times, each operation halving the feature map size and doubling the channel number. After the final downsampling process, the feature map size was 32×32 pixels, with the channel count reaching 1,024.

Transition layer

In this study, the PAM was incorporated into the transition layer. Following the final downsampling process, the feature map underwent two convolution operations for further extraction of the high-level features, with a consistent channel count of 1,024 being maintained. Subsequently, it was entered into the PAM for attention-information extraction. The detailed structure of the PAM is shown in *Figure 2*.

In this process, the input feature map undergoes a convolution operation complemented by a batch normalization layer and a nonlinear activation function,

resulting in two novel feature maps, X and Y. These maps, $\{X,Y\} \in R^{C \times H \times W}$, are reshaped into $\{X',Y'\} \in R^{C \times M}$, where $M = H \times W$. After this, matrix multiplication is applied to X' (after transposition) and Y'. The resultant spatial attention feature map $T \in R^{M \times M}$ can be obtained after passage through a softmax activation layer, as follows:

$$T_{ik} = \frac{\exp(X_i \cdot Y_k)}{\sum_{l=1}^{M} \exp(X_i \cdot Y_i)} \qquad [1]$$

where T_{ik} denotes the influence of the ith position on the kth position. A higher similarity between the features of the two positions results in a stronger impact on T_{ik}. $\sum_{l=1}^{M}$ denotes the summation that runs over all positions from $l=1$ to M, where M denotes the total number of positions. The denominator normalizes these scores across all positions, ensuring that the scores are in the [0,1] range and sum to 1.

Moreover, the feature map Z is generated by feeding the feature map Q into a convolution layer equipped with batch normalization and a nonlinear activation function, producing $Z \in R^{C \times H \times W}$. This is further reshaped to $Z' \in R^{C \times M}$.
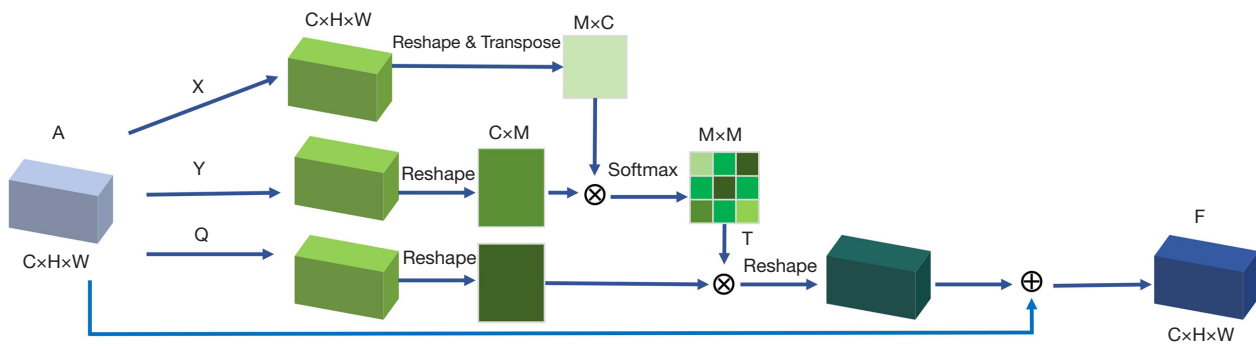
Figure 2 Architecture of the position attention module. A, input feature map; C, number of channels in the feature map; H, height of the feature map; W, width of the feature map; X, new feature map obtained after convolution operation, batch normalization layer, and nonlinear activation function; Y, new feature map obtained after convolution operation, batch normalization layer, and nonlinear activation function; Q, new feature map obtained after convolution operation, batch normalization layer, and nonlinear activation function; M, the size of the resulting matrix after transformation, equivalent to Height times Width (H×W); T, attention map generated after the softmax operation on the matrix product of X' (transposed) and Y'; F, final output feature map after the attention.

After matrix multiplication of Z' with T, the output is reshaped back to $R^{C \times H \times W}$. After the result is multiplied with factor β, it is element-wise added to Q to procure the final output $F \in R^{C \times H \times W}$, as follows:

$$F_k = \beta \sum_{i=1}^{M} \left(T_{ik} Z_i + Q_k \right) \qquad [2]$$

where β is initialized to zero and gradually learned throughout the training process.

From Eq. [2], it is evident that each position in F is an aggregation of the weighted sum of the features from all positions combined with the original feature. This ensures that the output captures comprehensive semantic information.

Upsampling section

The feature map obtained from the transition layer was subjected to a transposed convolution operation for upsampling. The image size increased from 32×32 to 64×64 pixels, and the number of channels decreased from 1,024 to 512. This was then concatenated along the channel dimensions with the corresponding feature map from the downsampling path. After concatenation, the number of channels was 1,024. Following the two convolution and activation operations, the number of channels was reduced to 512. This process was repeated four times; with each iteration, the feature map size doubled and the number of channels was halved. Upon completion of the final upsampling operation, the feature map size was 512×512 pixels. After passage through the output convolution layer, there

were two channels.

Loss function

Given the nature of this study, which focused on the automated segmentation of the GTV, we opted for cross-entropy loss as our loss function due to its efficacy in capturing pixel-wise discrepancies between the predicted outcomes and actual labels. This loss quantifies how well the predicted probability distribution matches the true distribution. For binary classification tasks that distinguish between tumor and nontumor regions, the mathematical representation of the cross-entropy loss can be expressed as follows:

$$L_{CE} = -\sum_{i=1}^{N} \left[y_i \log \left(p_i \right) + \left(1 - y_i \right) \log \left(1 - p_i \right) \right] \qquad [3]$$

where N denotes the total number of pixels in the image; y_i denotes the ground truth label of the ith pixel; which takes a value of 1 if the pixel belongs to the tumor region and 0 otherwise; and p_i denotes the predicted probability that the ith pixel belongs to the GTV region. The sum runs over all the pixels in the image, and the loss value provides a measure of the dissimilarity between the predicted and true labels.

Model training and validation

This study was conducted using the PyTorch framework with the experimental environment set up on an AMD R7 5950X CPU and an RTX 4090 24G GPU. The maximum

number of training epochs was set to 1,000. An early-stopping strategy was used to prevent overfitting. The Adam optimizer was used with a batch size of 16, an initial learning rate of 0.0001, and a learning rate decay of 10^{-6}. The dataset from The Second Affiliated Hospital of Nanchang University (n=123) was used for model training, whereas the independent dataset from The Affiliated Hospital of Southwest Medical University (n=45) served as the external validation dataset.

Model evaluation

In assessing the performance of the proposed model in the segmentation of the GTV, we used multiple metrics to provide a comprehensive evaluation of segmentation quality and accuracy. These metrics are described below.

(I)   Dice similarity coefficient (DSC): the DSC is a widely adopted metric for image segmentation tasks, particularly for medical imaging (22). It is a measure of the set similarity commonly used to compute the similarity between two masks. It can be defined as follows:

$$DSC = \frac{2 \times |P \cap G|}{|P| + |G|} \qquad [4]$$

where P denotes the predicted segmentation, and G denotes the ground truth.

(II) Intersection over union (IoU): the IoU measures the overlap between the predicted segmentation and the ground truth (23). It can be expressed as follows:

$$IoU = \frac{|P \cap G|}{|P \cup G|} \qquad [5]$$

(III) Accuracy: this is a straightforward metric that calculates the ratio of correctly predicted pixels to the total number of pixels (24), as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad [6]$$

where TP, TN, FP, and FN denote the true positives, true negatives, false positives, and false negatives, respectively.

(IV) Matthews correlation coefficient (MCC): the MCC is a metric that provides insight into the quality of binary classification. It returns a value between –1 and 1, where 1 indicates a perfect prediction, 0 indicates a random prediction, and –1 indicates an inverse prediction (25). It can be defined as follows:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \qquad [7]$$

(V) Hausdorff distance (HD): the HD measures the extent to which each point in a segmented image can be closely matched by a point in the ground truth, and vice versa. This is a measure of the largest of all directed distances from one point in one set to the closest point in the other set (26). Mathematically, given two nonempty sets P (the predicted segmentation) and G (the ground truth), the directed distance from set P to set G can be defined as follows:

$$h(P,G) = \max_{p \in P} \min_{g \in G} d(p,g) \qquad [8]$$

where $d(p,g)$ denotes the Euclidean distance between points p and g.

The HD can then be expressed as follows:

$$HD(P,G) = \max\{h(P,G), h(G,P)\} \qquad [9]$$

The results of conventional U-Net automatic segmentation on the external validation dataset and those of the PAM U-Net model on the external validation dataset were evaluated using the above-described evaluation metrics.

In this study, the gradient-weighted class activation mapping (Grad-CAM) algorithm was used to visualize the contribution distribution of CNN prediction outputs for selected samples. The Grad-CAM algorithm calculates the weight of each feature map in the last convolutional layer relative to the image class. The weighted sum of each feature map is then computed, and finally, the weighted sum feature map is mapped back to the original image. The Grad-CAM algorithm can be represented by the following formula:

$$L^c_{Grad-CAM} \in R^{u \times v} = ReLU\left(\sum_k \alpha^c_k A^k\right) \qquad [10]$$

Where $L^c_{Grad-CAM}$ is the Grad-CAM heatmap for class c; α^c_k represents the weight of the k-th feature map for class c, calculated as the global average-pooling for the gradients of the score for class c with respect to the feature map A^k; and A^k is the k-th feature map of the last convolutional layer.

Reader study

Fifty samples from the internal validation set were randomly selected for a prospective reader study. Five

Table 1 Automatic segmentation evaluation metrics

Evaluation metric	PAM U-Net	U-Net
Dice coefficient	0.753±0.172	0.691±0.142
Intersection over union	0.672±0.159	0.597±0.163
Accuracy	0.948±0.125	0.865±0.074
Sensitivity	0.721±0.116	0.669±0.127
Specificity	0.963±0.104	0.975±0.110
Matthews correlation coefficient	0.759±0.108	0.673±0.138
HD (mm)	6.9±0.6	8.2±0.8

All data are presented as mean ± standard deviation. PAM, position attention module; HD, Hausdorff distance.

radiation oncologists participated in the study, including one with 10 years of experience (radiation oncologist 1), one with 5 years of experience (radiation oncologist 2), one with 3 years of experience (radiation oncologist 3), and one with 1 year of experience (radiation oncologist 4). All radiation oncologists performed GTV delineation on the selected dataset without knowing the ground truth or the results of the automatic segmentation model. After a 4-week washout period, the segmentation results of PAM U-Net were provided to all participants as an auxiliary reference. The participants then resegmented the GTV of the selected dataset cases, integrating their judgment with the assistance provided by the PAM U-Net. Finally, the Dice coefficients from both segmentation rounds were compared to evaluate the applicability of PAM U-Net. Additionally, in order to compare segmentation results under different assistance conditions and to verify whether the proposed model provided meaningful enhancement, we recruited another radiation oncologist with 1 year of clinical experience (radiation oncologist 5) to perform GTV delineation every 4 weeks after a washout under a unassisted condition, a standard U-Net assistance condition, and a PAM U-Net assistance condition.

Statistical analysis

The segmentation capabilities of the PAM-improved U-Net and standard U-Net models were assessed using the DSC, IoU, accuracy, MCC, and HD metrics. The metrics averaged across patients were analyzed using a two-tailed paired *t*-test. Paired *t*-test was used to compare between manual segmentation of oncologists with and

without the assistance of the proposed model. Additionally, the consistency between the manually annotated and algorithmically predicted volumes was ascertained using the Linear regression analysis.

Results

GTV segmentation performance

The study population consisted of 79 females (47%) and 89 males (53%), with a median age of 61 years [interquartile range (IQR) 54–69 years]. The overall segmentation performances of the PAM-improved U-Net and standard U-Net models on the GTV of patients with brain metastases are presented in *Table 1*. The patient-averaged Dice coefficient for the PAM-improved U-Net model was 0.753±0.172 (median 0.786, IQR 0.69–0.85). For the standard U-Net model, the average Dice coefficient was 0.691±0.142 (median 0.73, IQR 0.61–0.80). Moreover, the segmentation results exhibited a significant difference between the two groups (P<0.001). The average HD for the standard U-Net model was 8.2±0.8 mm (median 8.6, IQR 7.3–9.0 mm), whereas for the PAM-improved U-Net model, the average HD was 6.9±0.6 mm (median 7.5, IQR 5.9–7.7 mm), implying a significant improvement with the inclusion of the PAM (P<0.001). The slices from the automatic segmentation results of the proposed PAM-improved U-Net and the standard U-Net models were assembled to form three-dimensional (3D) volumes and analyzed for Linear regression analysis with the ground truth (manual volume). The Linear regression coefficients for the PAM-improved U-Net and standard U-Net models were 0.926 and 0.874, respectively (both P values <0.001), demonstrating that the segmentation results from the PAM-improved U-Net model exhibited a better correlation with the manual volume (*Figure 3*).

Qualitative performance

In this study, three samples from the external validation dataset were randomly selected to observe the quality of the automatic segmentation masks and the qualitative differences between the models for some patients (*Figure 4*). For samples *Figure 4A-4C*, the DSC values of the standard U-Net automatic segmentation results were 0.625, 0.690, and 0.608, respectively, whereas those for the PAM-improved U-Net automatic segmentation results were 0.716, 0.747, and 0.782, respectively. It is evident that the
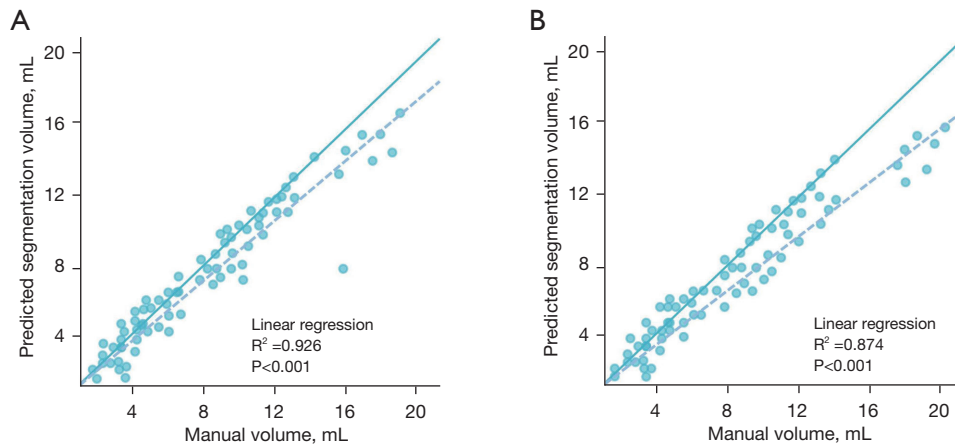
Figure 3 Scatter plot of the Linear regression analysis between the two models and the manual volume. (A) PAM-improved U-Net model. (B) Standard U-Net model. The solid line represents the ideal scenario where predicted volumes perfectly match the manual volumes. The dashed line represents the linear regression fit of the data points, indicating the actual relationship between the predicted and manual volumes. PAM, position attention module.
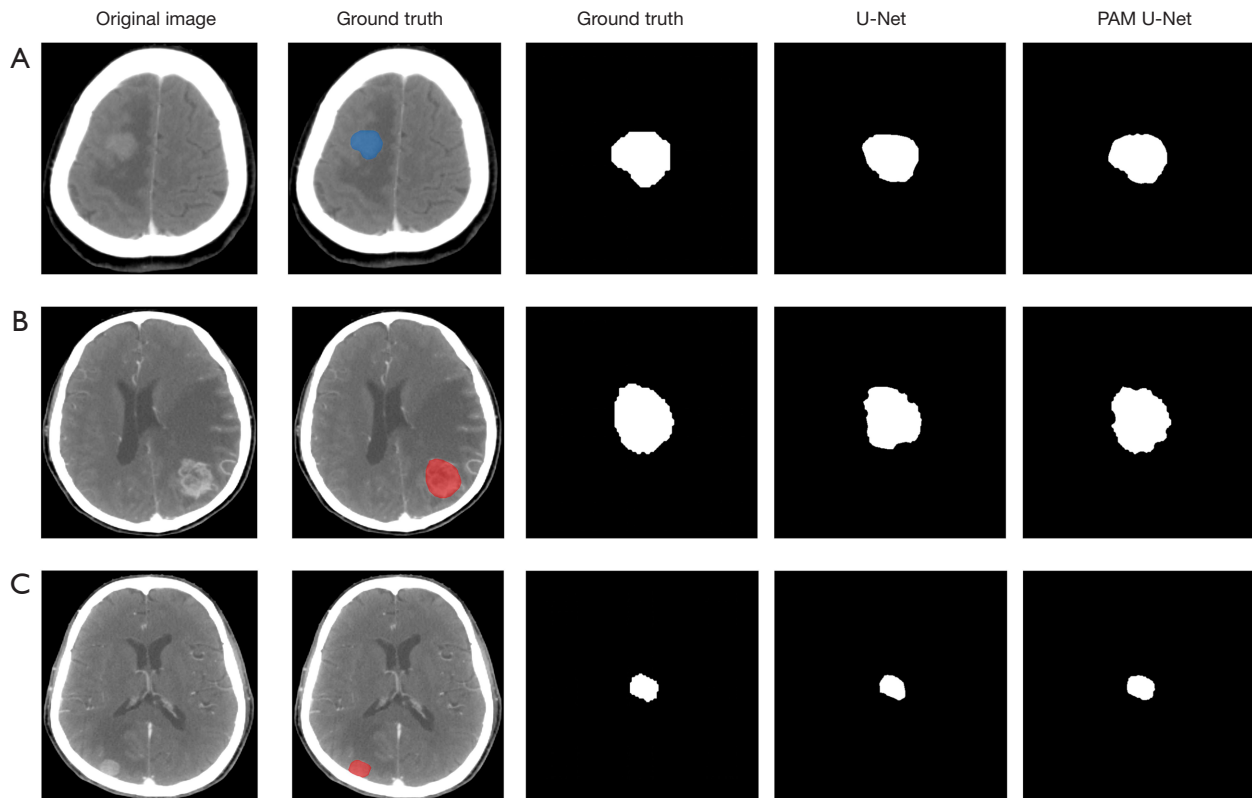

Figure 4 Results mask of automatic segmentation. Sample (A) is from the internal validation set, and (B,C) are from the external validation set. The blue and red regions are the ground truth represented in the original image. PAM, position attention module.
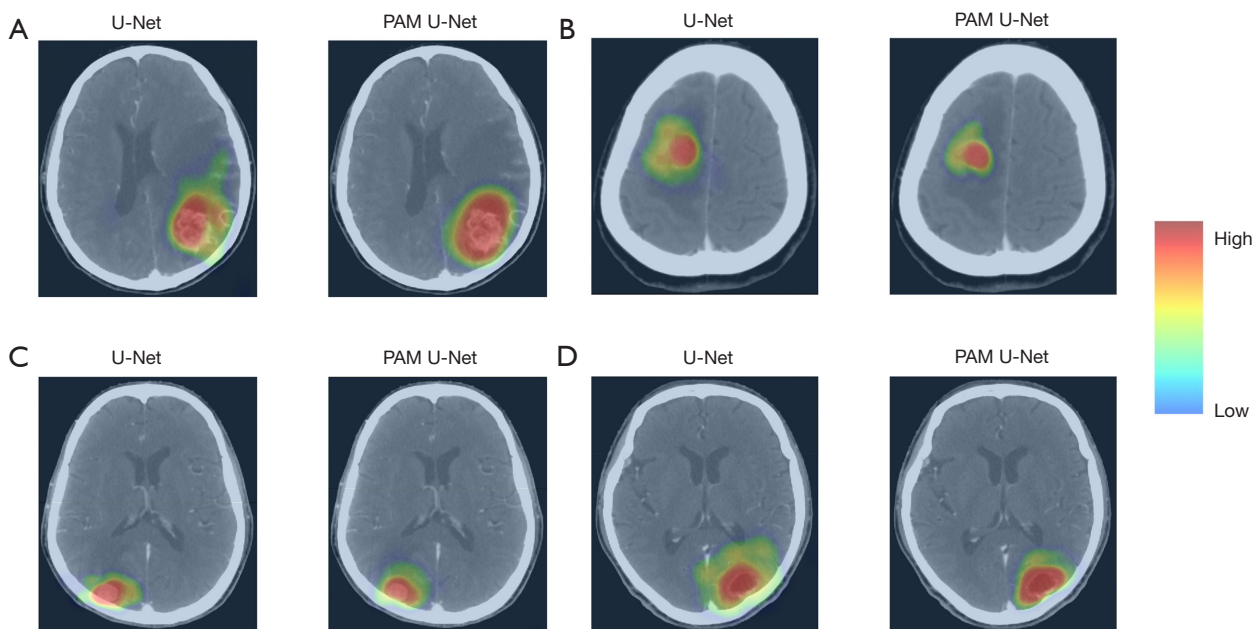
Figure 5 Visualization of the predictive outputs of the convolutional neural networks. PAM, position attention module.

automatic segmentation results of the PAM-improved U-Net model for smaller brain metastases are closer to the ground truth, with improvements from the U-Net model incorporating the PAM being more pronounced.

We also employed Grad-CAM algorithm to visualize the contribution distribution of the predictive outputs of the CNNs (*Figure 5*). In the Grad-CAM attention maps, the PAM U-Net model exhibited distinctly concentrated hotspots (notably marked by red regions) in contrast to the standard U-Net. These focal areas signify the regions where PAM U-Net allocated heightened attention, indicative of their perceived importance in the prediction process. Conversely, the standard U-Net displayed a more diffuse pattern of attention, suggesting a less targeted area of interest. Such a pattern might reflect a diminished precision in tumor localization. The attention maps for PAM U-Net demonstrated an increased localization surrounding the tumor areas. This focused attention is emblematic of the model's enhanced sensitivity to tumor-specific features, a testament to the efficacy of the PAM in guiding the network toward the salient spatial characteristics essential for accurate segmentation. Further scrutiny of the Grad-CAM visualizations revealed that PAM U-Net consistently highlighted the tumor regions with greater intensity and precision compared to U-Net. This observation is indicative of the model's sophisticated ability to discern and

emphasize pertinent tumor regions, which is paramount for segmentation and radiotherapy planning. These findings not only underscore the augmented capability of PAM U-Net in accurately identifying and segmenting tumor areas but also demonstrate the practical applicability of attention mechanisms in improving the interpretability and performance of CNN models in medical imaging tasks. However, it should be noted that the Grad-CAM attentional map has not been qualitatively studied at present and can only be judged by intuitive visual observation. Therefore, the Grad-CAM results merely represents a supplementary explanation for the quantitative analysis results.

Prospect reader study: the effect of automated segmentation model assistance on the segmentation results of the radiation oncologists

Figure 6 illustrates the comparison of the Dice coefficients between manual segmentation with and without the proposed model assistance. The experienced oncologists showed a nonsignificant (ns) change in performance, suggesting their expertise may enable them to achieve high segmentation accuracy without assistance. However, a statistically significant improvement (*P<0.05, ***P<0.001) was observed with the less experienced oncologists, highlighting the PAM U-Net's utility in enhancing
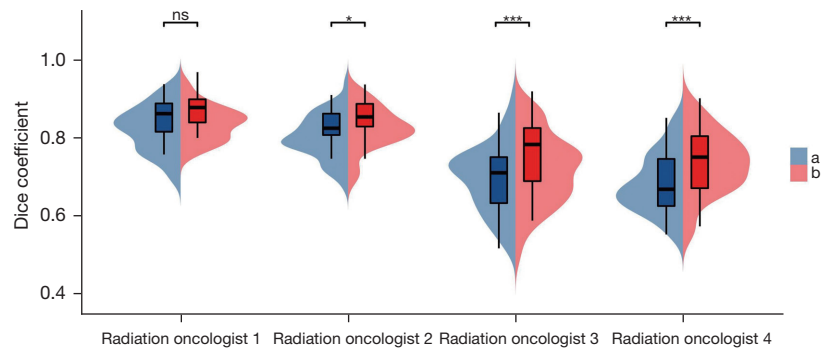
Figure 6 Comparison of Dice coefficients for manual segmentation with and without assistance from the proposed model performed by radiation oncologist 1 (10 years of experience), radiation oncologist 2 (5 years of experience), radiation oncologist 3 (3 years of experience), and radiation oncologist 4 (1 years of experience). Group a represents segmentation without the use of the proposed model for assisted segmentation, and group b represents segmentation with the use of the proposed model for assisted segmentation. *, P<0.05; ***, P<0.001. ns, not statistically significant.

Table 2 Segmentation results of radiation oncologist 5 (1 year of experience) under different conditions of segmentation model assistance

Metric	PAM U-Net-assisted	Standard U-Net-assisted	Unassisted
DSC	0.767±0.125	0.701±0.180	0.593±0.222
IoU	0.681±0.127	0.602±0.172	0.518±0.195
Accuracy	0.955±0.119	0.881±0.150	0.699±0.207
Sensitivity	0.741±0.115	0.680±0.168	0.611±0.215
Specificity	0.965±0.130	0.968±0.183	0.806±0.180
MCC	0.773±0.122	0.714±0.161	0.598±0.193
HD (mm)	8.8±1.6	10.2±4.8	17.5±6.9

All data are presented as the mean ± standard deviation. DSC, Dice similarity coefficient; IoU, intersection over union; MCC, Matthews correlation coefficient; HD, Hausdorff distance.

segmentation precision for less experienced practitioners. *Table 2* depicts the segmentation performance of radiation oncologist 5 under unassisted conditions, standard U-Net assistance, and PAM U-Net assistance. The improved metrics with the PAM U-Net assistance across all measured parameters—DSC, IoU, accuracy, sensitivity, specificity, MCC, and HD—attest to the model's effectiveness. Notably, the substantial reduction in HD when assisted by PAM U-Net reflects a more precise alignment with the ground truth, indicating a meaningful clinical impact on segmentation accuracy. *Figure 7* provides further evidence of the automated segmentation model's impact, displaying the boxplots of Dice coefficients under unassisted, standard U-Net-assisted, and PAM U-Net-assisted scenarios of radiation oncologist 5 (1 year of experience). PAM U-Net assistance produced higher median DSC values

(0.767±0.125) and narrower quartile spacing, indicating a statistically significant improvement in the segmentation effect compared to unassisted and standard U-Net-assisted descriptions (***P<0.001).

Comparative analysis with other U-Net variants

The evaluation metrics in *Table 3* offer insight into the comparative analysis with various U-Net variants, each with specific modifications. The DSC of PAM U-Net's indicates its effectiveness in segmentation, slightly outperforming the Swin Transformer U-Net (ST-U-Net) and showing improvement compared to residual U-Net (ResU-Net). This progress is clinically relevant, as it can potentially increase the segmentation reliability for treatment planning. The IoU metric achieved by PAM U-Net supports its

capability to precisely segment tumor boundaries, a critical attribute for ensuring accuracy in medical applications. In terms of overall accuracy, PAM U-Net exhibited a superior ability in differentiating between tumor and nontumor regions, which is crucial for radiotherapy planning. Although the specificity of PAM U-Net was marginally lower than the squeeze-and-excitation and attention module U-NET (SEA-U-Net), it effectively identified
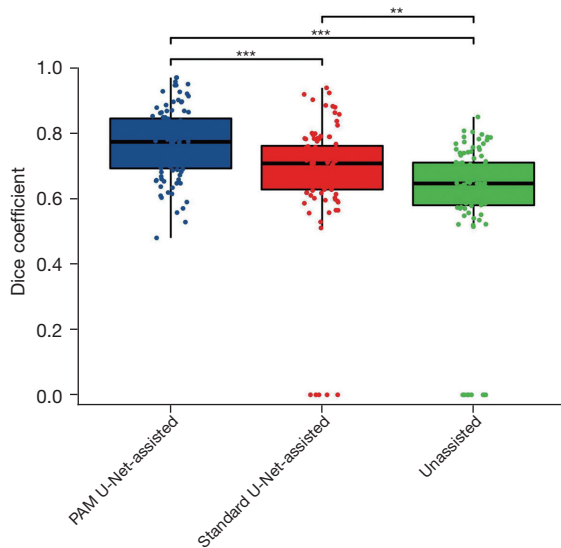
true negative cases, essential for minimizing overtreatment risks. The sensitivity of PAM U-Net was slightly below that of ST-U-Net, indicating areas where ST-U-Net may have a marginal advantage in detecting tumors. The MCC confirmed PAM U-Net as a robust segmentation tool, although it ranks just below ST-U-Net in this metric. Taken together, the evaluated metrics confirm PAM U-Net to be a comprehensive model capable of effectively managing binary classification challenges in medical imaging. Lastly, PAM U-Net yielded the lowest HD among the compared models, suggesting its segmentation contours align closely with actual tumor margins, an essential factor for precision in treatment planning and execution.

Discussion

In this study, the PAM-improved U-Net model demonstrated promising automatic segmentation performance on GTV segmentation tasks for patients with brain metastasis on an external validation dataset, achieving a DSC of 0.753 ± 0.172. Moreover, the proposed automatic segmentation model outperformed the standard U-Net model in terms of the DSC, IoU, accuracy, sensitivity, and MCC metrics.

This study is the first to employ CNNs for the automatic segmentation of brain metastasis CT images. Previous research on the automatic segmentation of brain metastases has been conducted on MRI images. Cao *et al.* (27) adopted an asymmetric U-Net architecture for the automatic segmentation of T1 sequence MRI scans, achieving a DSC of 0.84 in internal validation. Hsu *et al.* (28)

Figure 7 Comparison boxplots of Dice coefficient results for GTV segmentation by radiation oncologist 5 in unassisted, standard U-Net-assisted, and PAM U-Net-assisted, conditions, respectively. **, P<0.01; ***, P<0.001. GTV, gross tumor volume; PAM, position attention module.

Table 3 Comparison of the U-Net variants

Metrics	PAM U-Net	ST-U-Net (20)	SEA-U-Net (19)	SERR-U-Net (18)	ResU-Net (17)
DSC	0.753±0.172	0.747±0.158	0.730±0.135	0.718±0.156	0.706±0.163
IoU	0.672±0.159	0.667±0.143	0.648±0.150	0.625±0.141	0.610±0.157
Accuracy	0.948±0.125	0.930±0.131	0.919±0.118	0.898±0.122	0.881±0.117
Sensitivity	0.721±0.116	0.749±0.120	0.702±0.131	0.694±0.126	0.690±0.123
Specificity	0.963±0.104	0.951±0.112	0.978±0.106	0.946±0.114	0.921±0.110
MCC	0.759±0.108	0.768±0.105	0.730±0.097	0.714±0.119	0.696±0.133
HD (mm)	6.9±0.6	7.2±0.8	7.5±0.5	7.7±0.7	8.1±0.9

All data are presented as the mean ± standard deviation. PAM U-Net, positional attention module U-Net; ST-U-Net, Swin Transformer U-Net; SEA-U-Net, squeeze-and-excitation and attention module U-Net; SERR-U-Net, squeeze-and-excitation residual and recurrent block-based U-Net; ResU-Net, residual U-Net; DSC, Dice similarity coefficient; IoU, intersection over union; MCC, Matthews correlation coefficient; HD, Hausdorff distance.

conducted research using the V-Net structure for automatic segmentation of contrast-enhanced T1 MRI scans. They reported a DSC of 0.97 on their training dataset and 0.76 on their validation dataset, providing a reference for automatic segmentation of brain metastases. Synthetic CT from MRI is emerging as a valuable tool for precise segmentation in clinical settings. A previous study examined the automatic segmentation of brain metastases by generating synthetic CT from MRI, yet its integration into radiotherapy planning faces hurdles due to the discrepancies in patient positioning and support systems between MRI and CT scans (29). Variations in positioning devices such as RT flat tabletops, thermoplastic masks, and coil setups during MRI and CT procedures can introduce anatomical shifts, which may crucially effect brain radiotherapy in which millimeter accuracy is vital for targeting metastases and safeguarding vital structures (30). Additionally, the specialized MRI simulation equipment needed for consistent positioning is not universally accessible, particularly in resource-constrained environments, restricting synthetic CT's widespread adoption. These obstacles need to be overcome in order for synthetic CT to have comparable reliability in radiotherapy to that of traditional CT, which continues to be the standard care of treatment planning (31). Enhancing CT-based automatic segmentation algorithms could thus ensure broader access to high-quality care and promote global equity in care.

This study used a PAM-improved U-Net model for the task of segmenting the GTV in CT simulation images for brain metastases, achieving a DSC of 0.753±0.172 on the external validation dataset. This was comparable to the automatic segmentation on the MRI GTV validation datasets results of Hsu *et al.* (28). Moreover, the proposed model was directly applied to CT simulation imaging tasks, eliminating potential errors that could arise when MRI scans are merged with CT simulation images for GTV segmentation. Consequently, the proposed model could offer greater clinical utility for GTV segmentation in radiation therapy simulations.

In this study, the PAM significantly enhanced the classical U-Net architecture. The PAM-improved U-Net model generally outperformed the standard U-Net model across most metrics, although the specificity score of the standard U-Net model was slightly higher (0.975±0.110) than that of the PAM-improved U-Net model (0.963±0.104). This suggests that while the PAM-improved U-Net model was more adept at identifying positive cases, the standard U-Net model proved marginally better at excluding negative cases.

The sensitivity score of the PAM-improved U-Net model exceeded that of the standard U-Net model, indicating its superior ability to correctly identify true-positive cases, which is particularly beneficial for the detection of brain metastases.

The underlying mechanism of the PAM involves the selective emphasis on spatial relationships within feature maps, enabling the capture of long-range dependencies without the need to rely on an increased receptive field, thereby ensuring that the model gathers global contextual information from different spatial positions (32,33). This attention mechanism aids in refining local features using more globally aggregated features (34,35). A key advantage of integrating the PAM is its ability to weigh the importance of various spatial positions differently, thereby enabling the network to focus on regions with higher contextual relevance (36,37). This leads to improved segmentation performance, especially in intricate medical images, where tumors and other features may differ based on their spatial context (38). In our study, the PAM was strategically incorporated into the final block of the U-Net architecture to enhance its segmentation capabilities. This decision was driven by a careful balance between model complexity and performance efficiency. While it is feasible to integrate PAM after each block, similar to the application of squeeze-and-excitation or Res-Net modules in other U-Net variants, our approach aimed to optimize the tradeoff between computational burden and segmentation accuracy. Incorporating PAM at every stage would significantly increase the number of parameters, leading to a more complex model that requires more computational resources and training time. Such an increase in model complexity might not necessarily translate into a proportional improvement in segmentation performance, especially given the inherent challenges of medical image analysis. Additionally, excessive complexity could lead to overfitting, particularly when limited training data are used, a common scenario in medical imaging studies. By positioning the PAM in the final block, we aimed to capture the most relevant spatial contextual information at a stage where the feature maps are already highly refined. This strategic placement allows the PAM to focus on enhancing the feature representation with a global perspective, thereby improving the model's ability to highlight crucial areas for segmentation without overwhelming the network with redundant computations. With minimal addition of parameters, the PAM demonstrated an efficient and effective method for enhancing the spatial discernment

abilities of CNNs. Accurate and precise tumor delineation is critical to effective radiotherapy planning, in which the goal is to maximize the radiation dose to the tumor while minimizing exposure to the surrounding healthy tissues. The enhanced segmentation accuracy provided by the PAM U-Net model could lead to more precise targeting of brain metastases, potentially improving treatment efficacy and reducing the risk of radiation-induced side effects. Furthermore, the reduction in segmentation variability and the need for manual corrections can significantly streamline the treatment-planning process, enhancing overall clinical workflow efficiency.

Although the model constructed in this study achieved promising results on an external independent validation set, further validation of the model generalization performance is needed. Although the integration of PAM into the last module of the U-Net architecture in this study can reduce the number of model parameters to improve operational efficiency, it still has the potential to reduce the performance of model segmentation, especially under conditions of detecting micro brain metastases. In the future, we will use multiple centers and larger sample sizes in order to validate the model's generalizability. In addition, since this study is the first conduct automatic segmentation with CT simulation images of brain metastases, there is still a lack of data from other algorithms used for similar segmentation of brain metastasis images with which to compare the results of this study. Nonetheless, we compared the segmentation performances of several mainstream variants of the U-Net algorithm. In future studies, we will further use state-of-the-art algorithms such as Vision Transformer for automatic segmentation of CT simulation images of brain metastases.

Conclusions

We developed an automatic segmentation model based on the PAM-improved U-Net architecture, designed for automatic segmentation of the GTV in CT simulation images of patients with brain metastasis. The model demonstrated promising performance on an external validation dataset, offering valuable automated segmentation references for radiation oncologists to delineate the GTV in CT simulation images.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-23-1627/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-23-1627/coif). All authors report that this study was supported by the Sichuan Provincial Medical Research Project Plan (No. S21004), the Gulin County People's Hospital-Affiliated Hospital of Southwest Medical University Science and Technology Strategic Cooperation Program (No. 2022GLXNYDFY05), the Sichuan Science and Technology Program (No. 2022YFS0616), the Key-Funded Project of the National College Student Innovation and Entrepreneurship Training Program (Nos. 202310632001, 202310632028, 202310632036, and No. 202310632093). The authors have no other 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. This study was approved by the ethics review committees of the Affiliated Hospital of Southwest Medical University (No. KY2023041) and Jiangxi Cancer Hospital (No. 2023KY082). Owing to the retrospective nature of the study, the requirement for informed consent was waived. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Achrol AS, Rennert RC, Anders C, Soffietti R, Ahluwalia MS, Nayak L, Peters S, Arvold ND, Harsh GR, Steeg PS, Chang SD. Brain metastases. Nat Rev Dis Primers 2019;5:5.

2. Corti C, Antonarelli G, Criscitiello C, Lin NU, Carey LA, Cortés J, Poortmans P, Curigliano G. Targeting brain metastases in breast cancer. Cancer Treat Rev 2022;103:102324.

3. Gondi V, Bauman G, Bradfield L, Burri SH, Cabrera AR, Cunningham DA, Eaton BR, Hattangadi-Gluth JA, Kim MM, Kotecha R, Kraemer L, Li J, Nagpal S, Rusthoven CG, Suh JH, Tomé WA, Wang TJC, Zimmer AS, Ziu M, Brown PD. Radiation Therapy for Brain Metastases: An ASTRO Clinical Practice Guideline. Pract Radiat Oncol 2022;12:265-82.

4. Lehrer EJ, Jones BM, Dickstein DR, Green S, Germano IM, Palmer JD, Laack N, Brown PD, Gondi V, Wefel JS, Sheehan JP, Trifiletti DM. The Cognitive Effects of Radiotherapy for Brain Metastases. Front Oncol 2022;12:893264.

5. Little MP, Patel A, Lee C, Hauptmann M, Berrington de Gonzalez A, Albert P. Impact of Reverse Causation on Estimates of Cancer Risk Associated With Radiation Exposure From Computerized Tomography: A Simulation Study Modeled on Brain Cancer. Am J Epidemiol 2022;191:173-81.

6. Schiff JP, Price AT, Stowe HB, Laugeman E, Chin RI, Hatscher C, Pryser E, Cai B, Hugo GD, Kim H, Badiyan SN, Robinson CG, Henke LE. Simulated computed tomography-guided stereotactic adaptive radiotherapy (CT-STAR) for the treatment of locally advanced pancreatic cancer. Radiother Oncol 2022;175:144-51.

7. Nelissen KJ, Versteijne E, Senan S, Rijksen B, Admiraal M, Visser J, Barink S, de la Fuente AL, Hoffmans D, Slotman BJ, Verbakel WFAR. Same-day adaptive palliative radiotherapy without prior CT simulation: Early outcomes in the FAST-METS study. Radiother Oncol 2023;182:109538.

8. Chapman JW, Lam D, Cai B, Hugo GD. Robustness and reproducibility of an artificial intelligence-assisted online segmentation and adaptive planning process for online adaptive radiation therapy. J Appl Clin Med Phys 2022;23:e13702.

9. Yu C, Anakwenze CP, Zhao Y, Martin RM, Ludmir EB, S Niedzielski J, Qureshi A, Das P, Holliday EB, Raldow AC, Nguyen CM, Mumme RP, Netherton TJ, Rhee DJ, Gay SS, Yang J, Court LE, Cardenas CE. Multi-organ segmentation of abdominal structures from non-contrast and contrast enhanced CT images. Sci Rep 2022;12:19093.

10. Abbani N, Baudier T, Rit S, Franco FD, Okoli F, Jaouen V, Tilquin F, Barateau A, Simon A, de Crevoisier R, Bert J, Sarrut D. Deep learning-based segmentation in prostate radiation therapy using Monte Carlo simulated cone-beam computed tomography. Med Phys 2022;49:6930-44.

11. Huang Y, Bert C, Sommer P, Frey B, Gaipl U, Distel LV, Weissmann T, Uder M, Schmidt MA, Dörfler A, Maier A, Fietkau R, Putz F. Deep learning for brain metastasis detection and segmentation in longitudinal MRI data. Med Phys 2022;49:5773-86.

12. Zhao JY, Cao Q, Chen J, Chen W, Du SY, Yu J, Zeng YM, Wang SM, Peng JY, You C, Xu JG, Wang XY. Development and validation of a fully automatic tissue delineation model for brain metastasis using a deep neural network. Quant Imaging Med Surg 2023;13:6724-34.

13. Papp J, Simon M, Csiki E, Kovács Á. CBCT Verification of SRT for Patients With Brain Metastases. Front Oncol 2021;11:745140.

14. Li B, Huang J, Ruan J, Peng Q, Huang S, Li Y, Li F. Dosimetric impact of CT metal artifact reduction for spinal implants in stereotactic body radiotherapy planning. Quant Imaging Med Surg 2023;13:8290-302.

15. Yang Y, Cao S, Wan W, Huang S. Multi-modal medical image super-resolution fusion based on detail enhancement and weighted local energy deviation. Biomed Signal Process Control 2023;80:104387.

16. Zhang J, Yang L, Hu Y, Leng X, Huang W, Liu Y, Liu X, Wang L, Zhang J, Li D, Tang L, Xiang J, Du C. Calculation of left ventricular ejection fraction using an 8-layer residual U-Net with deep supervision based on cardiac CT angiography images versus echocardiography: a comparative study. Quant Imaging Med Surg 2023;13:5852-62.

17. Lv P, Wang J, Zhang X, Ji C, Zhou L, Wang H. An improved residual U-Net with morphological-based loss function for automatic liver segmentation in computed tomography. Math Biosci Eng 2022;19:1426-47.

18. Wang J, Li X, Lv P, Shi C. SERR-U-Net: Squeeze-and-Excitation Residual and Recurrent Block-Based U-Net for Automatic Vessel Segmentation in Retinal Image. Comput Math Methods Med 2021;2021:5976097.

19. Xiong L, Yi C, Xiong Q, Jiang S. SEA-NET: medical

image segmentation network based on spiral squeeze-and-excitation and attention modules. BMC Med Imaging 2024;24:17.

20. Fang K, He B, Liu L, Hu H, Fang C, Huang X, Jia F. UMRFormer-net: a three-dimensional U-shaped pancreas segmentation method based on a double-layer bridged transformer network. Quant Imaging Med Surg 2023;13:1619-30.

21. Gong Y, Gu Z, Zhang Z, Ma L. CPSAM: Channel and Position Squeeze Attention Module. In: Mantoro T, Lee M, Ayu MA, Wong KW, Hidayanto AN. editors. Neural Information Processing. ICONIP 2021. Lecture Notes in Computer Science(), Springer, 2021:190-202.

22. Cardenas CE, McCarroll RE, Court LE, Elgohari BA, Elhalawani H, Fuller CD, Kamal MJ, Meheissen MAM, Mohamed ASR, Rao A, Williams B, Wong A, Yang J, Aristophanous M. Deep Learning Algorithm for Auto-Delineation of High-Risk Oropharyngeal Clinical Target Volumes With Built-In Dice Similarity Coefficient Parameter Optimization Function. Int J Radiat Oncol Biol Phys 2018;101:468-78.

23. Nowozin S. Optimal decisions from probabilistic models: the intersection-over-union case. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014:548-55.

24. Murphy KR, Garcia M, Kerkar S, Martin C, Balzer WK. Relationship between observational accuracy and accuracy in evaluating performance. J Appl Psychol 1982;67:320.

25. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 2020;21:6.

26. Zhao C, Shi W, Deng Y. A new Hausdorff distance for image matching. Pattern Recognit Lett 2005;26:581-6.

27. Cao Y, Vassantachart A, Ye JC, Yu C, Ruan D, Sheng K, Lao Y, Shen ZL, Balik S, Bian S, Zada G, Shiu A, Chang EL, Yang W. Automatic detection and segmentation of multiple brain metastases on magnetic resonance image using asymmetric UNet architecture. Phys Med Biol 2021;66:015003.

28. Hsu DG, Ballangrud Å, Shamseddine A, Deasy JO, Veeraraghavan H, Cervino L, Beal K, Aristophanous M. Automatic segmentation of brain metastases using T1 magnetic resonance and computed tomography images. Phys Med Biol 2021;66:10.1088/1361-6560/ac1835.

29. Putz F, Mengling V, Perrin R, Masitho S, Weissmann T, Rösch J, Bäuerle T, Janka R, Cavallaro A, Uder M, Amarteifio P, Doussin S, Schmidt MA, Dörfler A, Semrau S, Lettmaier S, Fietkau R, Bert C. Magnetic resonance imaging for brain stereotactic radiotherapy : A review of requirements and pitfalls. Strahlenther Onkol 2020;196:444-56.

30. Mekiš V, Žager Marciuš V, Rogina D, Dolenc L, Mekiš N. Comparison of treatment position with mask immobilization and standard diagnostic setup in intracranial MRI radiotherapy simulation. Strahlenther Onkol 2021;197:614-21.

31. Masitho S, Grigo J, Brandt T, Lambrecht U, Szkitsak J, Weiss A, Fietkau R, Putz F, Bert C. Synthetic CTs for MRI-only brain RT treatment: integration of immobilization systems. Strahlenther Onkol 2023;199:739-48.

32. Zhang Z, Xue H, Zhou G. A plug-and-play attention module for CT-based COVID-19 segmentation. J Phys: Conf Ser 2021. doi: 10.1088/1742-6596/2078/1/012041.

33. Shi Q, Tang X, Yang T, Liu R, Zhang L. Hyperspectral image denoising using a 3-D attention denoising network. IEEE Transactions on Geoscience and Remote Sensing 2021;59:10348-63.

34. Zhu Y, Chen C, Yan G, Guo Y, Dong Y. AR-Net: Adaptive attention and residual refinement network for copy-move forgery detection. IEEE Transactions on Industrial Informatics 2020;16:6714-23.

35. Xia Z, Pan X, Song S, Li LE, Huang G. Vision transformer with deformable attention. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022:4794-803.

36. Dong Y, Liu Q, Du B, Zhang L. Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification. IEEE Trans Image Process 2022;31:1559-72.

37. Fan T, Wang G, Li Y, Wang H. Ma-net: A multi-scale attention network for liver and tumor segmentation. IEEE Access 2020;8:179656-65.

38. Gu R, Wang G, Song T, Huang R, Aertsen M, Deprest J, Ourselin S, Vercauteren T, Zhang S. CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation. IEEE Trans Med Imaging 2021;40:699-711.