Original Article

Modality preserving U-Net for segmentation of multimodal medical images

Bingxuan Wu[1], Fan Zhang[1], Liang Xu[2], Shuwei Shen[2], Pengfei Shao[1], Mingzhai Sun[2], Peng Liu[2], Peng Yao[3], Ronald X. Xu[2]

[1]Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei, China; [2]Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, China; [3]School of Microelectronics, University of Science and Technology of China, Hefei, China

Contributions: (I) Conception and design: B Wu, P Yao, P Liu; (II) Administrative support: P Shao, RX Xu; (III) Provision of study materials or patients: B Wu, P Yao, RX Xu; (IV) Collection and assembly of data: B Wu, L Xu; (V) Data analysis and interpretation: B Wu, F Zhang, S Shen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Peng Liu, PhD. Suzhou Institute for Advanced Research, University of Science and Technology of China, 188 Ren'ai Road, Suzhou 215123, China. Email: lpeng01@ustc.edu.cn; Peng Yao, PhD. University of Science and Technology of China, 96 Huangshan Road, Hefei 230026, China. Email: yaopeng@ustc.edu.cn; Ronald X. Xu, PhD. Suzhou Institute for Advanced Research, University of Science and Technology of China, 188 Ren'ai Road, Suzhou 215123, China. Email: xux@ustc.edu.cn.

Background: Recent advances in artificial intelligence and digital image processing have inspired the use of deep neural networks for segmentation tasks in multimodal medical imaging. Unlike natural images, multimodal medical images contain much richer information regarding different modal properties and therefore present more challenges for semantic segmentation. However, there is no report on systematic research that integrates multi-scaled and structured analysis of single-modal and multimodal medical images.

Methods: We propose a deep neural network, named as Modality Preserving U-Net (MPU-Net), for modality-preserving analysis and segmentation of medical targets from multimodal medical images. The proposed MPU-Net consists of a modality preservation encoder (MPE) module that preserves the feature independency among the modalities and a modality fusion decoder (MFD) module that performs a multiscale feature fusion analysis for each modality in order to provide a rich feature representation for the final task. The effectiveness of such a single-modal preservation and multimodal fusion feature extraction approach is verified by multimodal segmentation experiments and an ablation study using brain tumor and prostate datasets from Medical Segmentation Decathlon (MSD).

Results: The segmentation experiments demonstrated the superiority of MPU-Net over other methods in the segmentation tasks for multimodal medical images. In the brain tumor segmentation tasks, the Dice scores (DSCs) for the whole tumor (WT), the tumor core (TC) and the enhancing tumor (ET) regions were 89.42%, 86.92%, and 84.59%, respectively. In the meanwhile, the 95% Hausdorff distance (HD95) results were 3.530, 4.899 and 2.555, respectively. In the prostate segmentation tasks, the DSCs for the peripheral zone (PZ) and the transitional zone (TZ) of the prostate were 71.20% and 90.38%, respectively. In the meanwhile, the 95% HD95 results were 6.367 and 4.766, respectively. The ablation study showed that the combination of single-modal preservation and multimodal fusion methods improved the performance of multimodal medical image feature analysis.

Conclusions: In the segmentation tasks using brain tumor and prostate datasets, the MPU-Net method has achieved the improved performance in comparison with the conventional methods, indicating its potential application for other segmentation tasks in multimodal medical images.

Keywords: Multimodal medical image segmentation; semantic segmentation; artificial intelligence

Submitted Dec 09, 2022. Accepted for publication May 19, 2023. Published online Jun 14, 2023.
doi: 10.21037/qims-22-1367
View this article at: https://dx.doi.org/10.21037/qims-22-1367

Introduction

Image segmentation is an important step for reconstructing the anatomical structures of relevant tissues and organs as preoperative images are analyzed for precise surgical navigation or accurate diagnosis of diseases (1-3). Automated lesion segmentation can provide physicians with critical information about tumor volume, location and shape, highlighting the core tumor region and the entire tumor area. In the process of biomedical image analysis, image segmentation help to focus on pathology (4), track disease progression (5), and characterize anatomical structures and defects (6). Moreover, it facilitates timely diagnosis and effective treatment of neurological disorders such as Alzheimer's disease (7) and Parkinson's disease (8). In these applications, image segmentation helps to generate quantitative measurements (e.g., mask of the lesion) for subsequent tasks of diagnostic and treatment planning. Thus, automated and reliable segmentation techniques play a pivot role in clinical management of many diseases.

In the field of image analysis, many segmentation methods have been developed and implemented, such as the active contour model (9), the atlas-based registration (10), the fuzzy clustering (11), the superpixel method (12) and the graph-cut method (13). Convolutional neural networks (CNNs) (14) and other deep learning methods have also been widely used in automated medical image analysis (15-18). Ronneberger *et al.* presented a network called U-Net, which consisted of a contracting path to capture context and a symmetric expanding path that enabled precise localization (19). U-Net has been a great success in the field of biomedical image segmentation, and many U-Net-based architectures have been developed since then. Bakas *et al.* reviewed the work on brain tumor segmentation over the years in detail (20). Kamnitsas *et al.* explored Ensembles of Multiple Models and Architectures (EMMA) for brain tumor segmentation and won first place in the BraTS (Brain Tumor Segmentation Challenge) 2017 (21). Myronenko described a semantic segmentation network based on encoder-decoder architecture, and added a variational auto-encoder branch to reconstruct the input image itself in order to regularize the shared decoder and impose additional constraints on its layers (22). Myronenko's approach won first place in the BraTS 2018. Qin *et al.* proposed U^2Net, which was also a successful architecture based on the U-Net structure for salient object detection (23). The most distinctive feature of U^2Net was a two-level nested U-structure which was able to capture more contextual information from different scales. Isensee *et al.* developed nnU-Net, a deep-learning based segmentation method that automatically configures itself, including preprocessing, network architecture, training and post-processing for any new task (24). In Medical Segmentation Decathlon (MSD) challenge, nnU-Net achieved state-of-the-art performance on many tasks including against task-optimized networks (25). Hatamizadeh *et al.* introduced a novel architecture, dubbed as UNEt TRansformers (UNETR), that utilized a transformer as the encoder to learn sequence representations of the input volume and effectively capture the global multi-scale information, while also following the successful "U-shaped" network design for the encoder and decoder (26). Tang *et al.* proposed a new 3D transformer-based model, dubbed Swin UNEt TRansformers (Swin UNETR), with a hierarchical encoder for self-supervised pre-training, and demonstrated successful pre-training of the proposed model on 5,050 publicly available computed tomography (CT) images from various body organs (27). Liu *et al.* introduced embedding learned from Contrastive Language-Image Pre-training (CLIP) to segmentation models, dubbed the CLIP-Driven Universal Model, which could better segment 25 organs and 6 types of tumors by exploiting the semantic relationship between abdominal structures (28). The various U-Net-based image segmentation techniques are shown in *Table 1*.

Despite these advances, automated segmentation of multimodal medical images presents unique challenges. An example of a multimodal medical images is shown in *Figure 1*. Since the relationship between different modalities is highly nonlinear, appropriate modal fusion is crucial for successful image segmentation (29). However, for the multimodal image segmentation tasks that rely on specific modalities, extracting valid information from multiple modalities is difficult if these modalities are not exploited concurrently (30). In this regard, most of the multimodal image analysis techniques adopted the early fusion strategy where multiple modalities are fused at the first stage

Table 1 Various U-Net-based image segmentation techniques

Methods	First author	Institution	Characteristic	Task	Year
U-Net (19)	Ronneberger	University of Freiburg, Germany	Consists of a contracting path to capture context and a symmetric expanding path that enables precise localization	Biomedical image segmentation	2015
EMMA (21)	Kamnitsas	Imperial College London, UK	Aggregation of predictions from a wide range of methods	Brain tumor segmentation	2017
Encoder-Decoder-VAE (22)	Myronenko	NVIDIA, USA	A variational auto-encoder branch is added to reconstruct the input image	Brain tumor segmentation	2019
U^2Net (23)	Qin	University of Alberta, Canada	Two-level nested U-structure to capture more contextual information from different scales	Salient object detection	2020
nnU-Net (24)	Isensee	German Cancer Research Center, Germany	Automatically configures segmentation method, including preprocessing, network architecture, training and post-processing for any new task	Biomedical image segmentation	2021
UNETR (26)	Hatamizadeh	NVIDIA, USA	Utilizes a transformer as the encoder to learn sequence representations of the input volume and effectively capture the global multi-scale information	Medical image segmentation	2022
SwinUNETR (27)	Tang	NVIDIA, USA	A new 3D transformer-based model with a hierarchical encoder for self-supervised pre-training	Medical image segmentation	2022
Universal Model (28)	Liu	City University of Hong Kong, Hong Kong	Introduce embedding learned from CLIP to segmentation models	Medical image segmentation	2023

EMMA, Ensembles of Multiple Models and Architectures; VAE, Variational Autoencoder; UNETR, UNEt TRansformers; CLIP, Contrastive Language-Image Pre-training.

of feature analysis (31). In addition, late fusion and the strategies that combines early and late fusion have also been explored (32,33). In summary, the current research effort has focused on how to fuse the primary single-modal information into multimodal information. Of all the literature we reviewed, there is no related work that extracts single modality features, analyzes the single modalities, and fuses the single-modal information at different scales with multimodal information. The lack of structured cross-modal information fusion at different scales leads to the loss of rich modal information in the network and limits its effectiveness and application in multimodal medical image analysis. Therefore, it becomes crucial to find more efficient methods that concurrently integrate the analysis of different imaging modalities (34).

In the development of artificial intelligence image analysis, some unique approaches have achieved great success and attracted much attention from researchers. The best-known method for biomedical image segmentation is U-Net (19), as they have significant predictive performance and are widely used due to their flexible architecture. In the encoder of the U-Net architecture, the network learns deep features, and in the decoder, the network performs segmentation based on the learned features. There are skip connections between each layer of the encoder and decoder for better information transfer, making the segmentation more successful. Due to the good performance of U-Net, many studies on the segmentation of multimodal medical images were improved on the basis of U-Net (23,24,26,35,36). In addition to U-Net, Group convolution (37) has received much attention as a basic module for feature extraction. Group convolution differs from the commonly used image convolution in that a channel of the output features is affected by only a portion of the input channels instead of all the input channels, as shown in *Figure 2*. Therefore, group convolution has great potential in multimodal medical image analysis by extracting features while maintaining independence between modalities.

To address the above problem, we designed a structured single-modal preservation and multimodal fusion medical image segmentation network called MPU-Net based on vanilla U-Net and group convolution. The network consists of a modality preservation encoder (MPE) module and a modality fusion decoder (MFD) module. MPE enables multiscale feature extraction of multimodal images, and MFD generates a segmentation mask based on the features extracted by MPE. In the encoder, we use both group

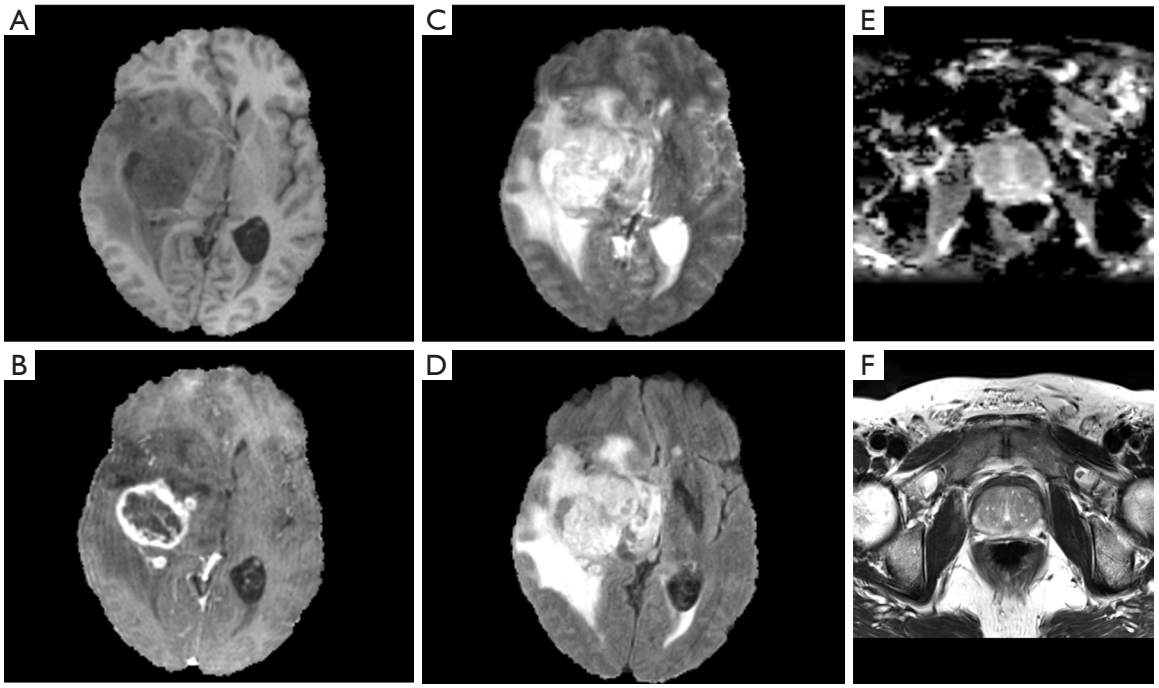

Figure 1 Representative multimodal medical MR images of brain tumor and the prostate. (A) Brain tumor in the T1 modality. (B) Brain tumor in the T1ce modality. (C) Brain tumor in the T2 modality. (D) Brain tumor in the FLAIR modality. (E) Prostate in the ADC modality. (F) Prostate in T2 modality. MR, magnetic resonance; T1, T1-weighted; T1ce, T1-weighted contrast-enhanced; T2, T2-weighted; FLAIR, fluid attenuated inversion recovery; ADC, apparent diffusion coefficient.

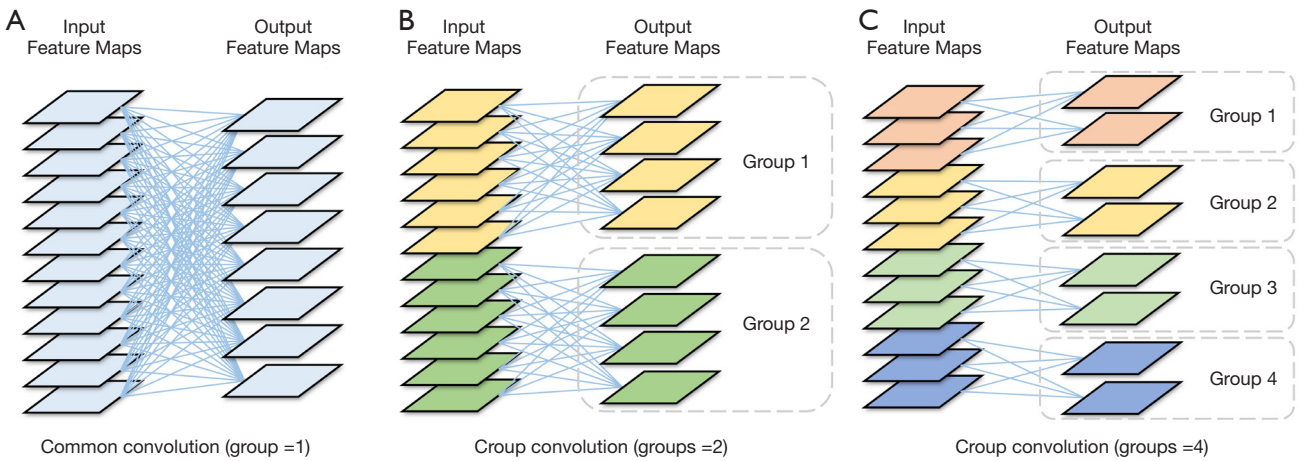

Figure 2 Demonstration of common convolution and group convolution. (A) Common convolution. (B) Group convolution with group of 2. (C) Group convolution with group of 4.

convolution and common convolution to achieve multiscale feature extraction with single-modality preservation and multimodal fusion, providing a basis for medical image segmentation tasks that may depend on specific modalities. In the decoder, we output a segmentation mask that combines the information of single-modality preservation and multimodal fusion, which can further improve the performance of medical image segmentation

tasks. Compared with other image segmentation networks (e.g., U-Net), we do not completely mix the different image modalities but retain the separate low-level and high-level semantic features of each modality, thus providing clearer features for modality-specific tasks as much as possible.

In this paper, we introduced MPU-Net and its components MPE and MFD and tested MPU-Net with a four-modality brain tumor image dataset and a two-modality prostate image dataset from MSD (25). We also compared MPU-Net with popular medical image segmentation methods to demonstrate the advantages of MPU-Net for multimodal medical image segmentation. Finally, we performed an ablation study to verify the effectiveness of modality fusion combined with modality preservation methods.

The main contributions of this paper are thus summarized as follows:

- ❖ We proposed a novel MPU-Net that was better adapted to multimodal medical image segmentation tasks by two innovative modules: the MPE module and the MFD module. In MPE, by using both group convolution and common convolution, we can obtain multimodal fused features while maintaining single-modal features. In MFD, the segmentation performance was improved by deeply fusing the multimodal features and single-modal features passed by MPE.
- ❖ We applied MPU-Net to two multimodal medical image segmentation databases, brain tumor and prostate. The comparison with other methods showed that our method achieved the best performance on the brain tumor dataset and second place on the prostate dataset. In addition, further ablation experiments demonstrated the advanced nature of our proposed MPE and MFD modules.

It is important to note that the impact of our study was not limited to multimodal brain tumor and prostate segmentation tasks. MPU-Net, as described in this paper, could be used as a base network and thus further extended to other multimodal medical image segmentation tasks. The source codes of our models are available at https://github.com/BinsonW/MPU-Net.

Methods

Network architecture

The proposed MPU-Net consists of two parts: an encoder and a decoder. We first introduced MPU-Net and then introduced the MPE, the basic component of the encoder, and the MFD, the basic component of the decoder.

MPU-Net

MPU-Net takes multimodal 3D images as input and outputs the corresponding segmentation mask, as shown in *Figure 3*. The modality preservation feature map and modality fusion feature map were first generated by performing common convolution and group convolution on 3D images, and then these two feature maps were input into the first MPE module. In the encoder part, MPU-Net extracted image features using 4 MPE modules to generate feature maps with different numbers of channels and spatial resolutions. In the decoder part, MPU-Net used 4 MFD modules to upsample the modality preservation feature map and modality fusion feature map, and output segmentation masks with different resolutions for deep supervision in the training phase or as output results.

The MPU-Net contained two paths for feature analysis. The first path used group convolution for single-modal feature extraction, and preserved the independence between different image modalities, providing a basis for tasks that may depend on specific modalities. The second path used common convolution for multimodal feature extraction, and fused modality-preserved and modality-fused information, thus ensuring that we could extract information from the fused modality information that was not available from a single modality.

To make full use of the different semantic information of single-modal and multimodal, multiple feature fusions were added to different scales and modalities within and between the two paths of feature analysis. Following U-Net's approach (19), we added a skip-connection between the encoder and decoder, as shown by the blue and orange dashed lines in *Figure 3*, which serves to retain more high-resolution detailed information embedded in the high-level feature maps, thus improving the image segmentation accuracy. The gray dashed lines in *Figure 3* indicate the concatenation and convolution of the modality preservation feature maps and modality fusion feature maps output in MFD, and the final segmentation mask was generated.

MPE

The MPE module, as shown in *Figure 4A*, was used in each layer of the encoder part. The MPE extracted features from the single-modal preservation and multimodal fusion data

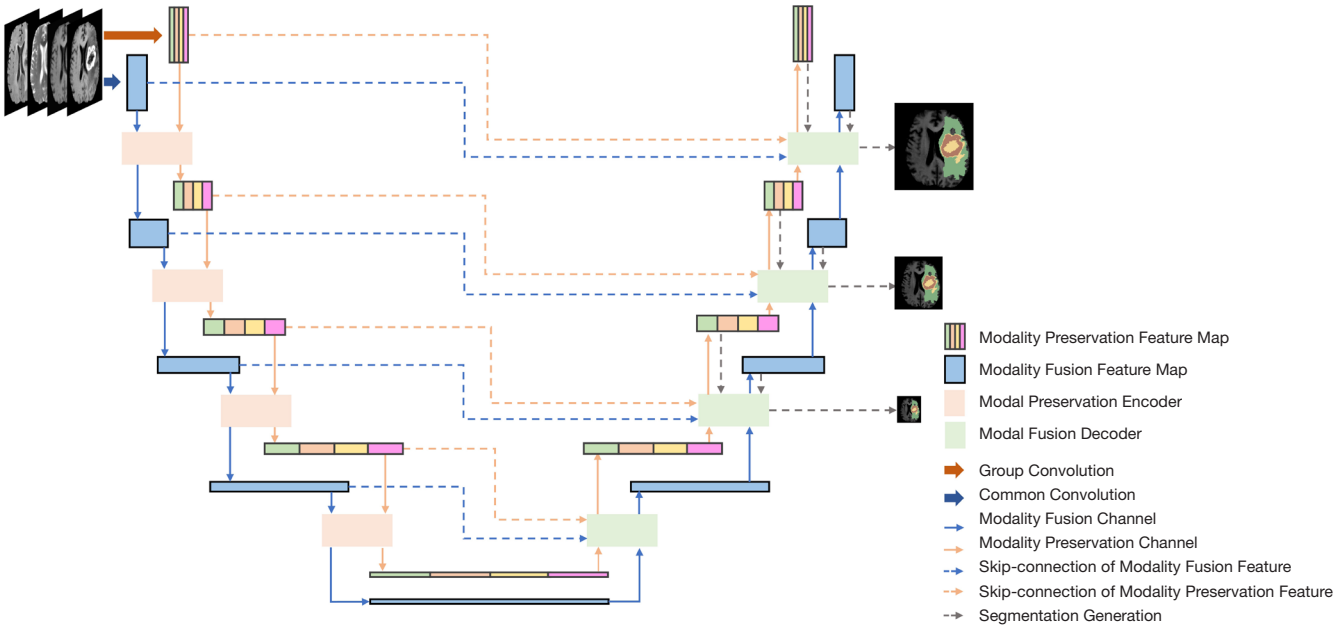

Figure 3 Diagram of the MPU-Net, which is composed of the encoder and decoder. The encoder and decoder contain 4 layers of the MPE and the MFD, respectively. The blue and orange dashed lines indicate the skip-connections for retaining more high-resolution detailed information, and the gray dashed lines indicate the skip-connections inter-modality. MPU-Net, Modality Preserving U-Net; MPE, modality preservation encoder; MFD, modality fusion decoder.

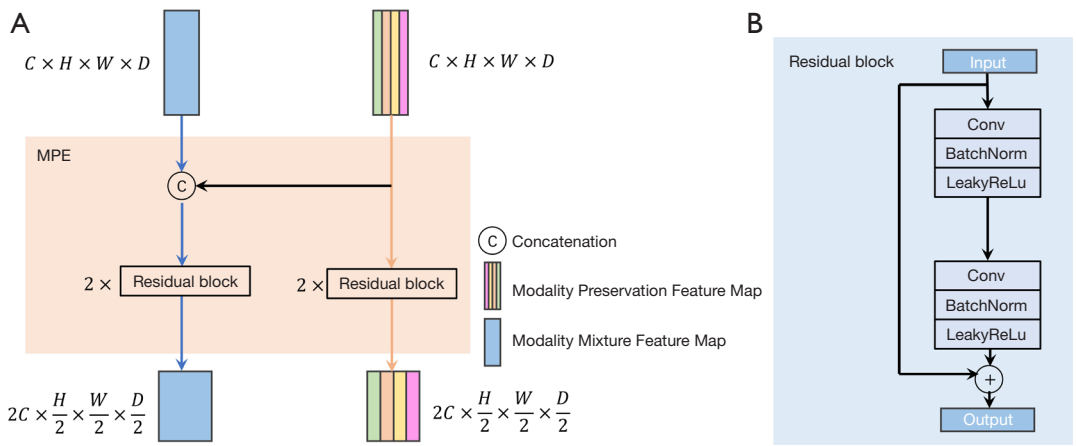

Figure 4 MPE and residual block. (A) The structure of MPE. C: channel; H: height; W: width; D: depth. (B) The structure of the residual block. Conv: convolution; BatchNorm: batch normalization. MPE, modality preservation encoder.

and output the single-modal preservation and multimodal fusion data as input to the next MPE module or the first MFD module. The MPE contained both single-modal preservation and multimodal fusion feature extraction paths, and the multimodal fusion features of each layer were fused with single-modal preservation features by concatenation (black line in *Figure 4A*). By superimposing the features of a single modality onto the multimodal features together, we believe that we can preserve the unique information of each modality in the single-modal channel (which unique information may have been lost in the multimodal fusion channel), provide useful information for feature extraction

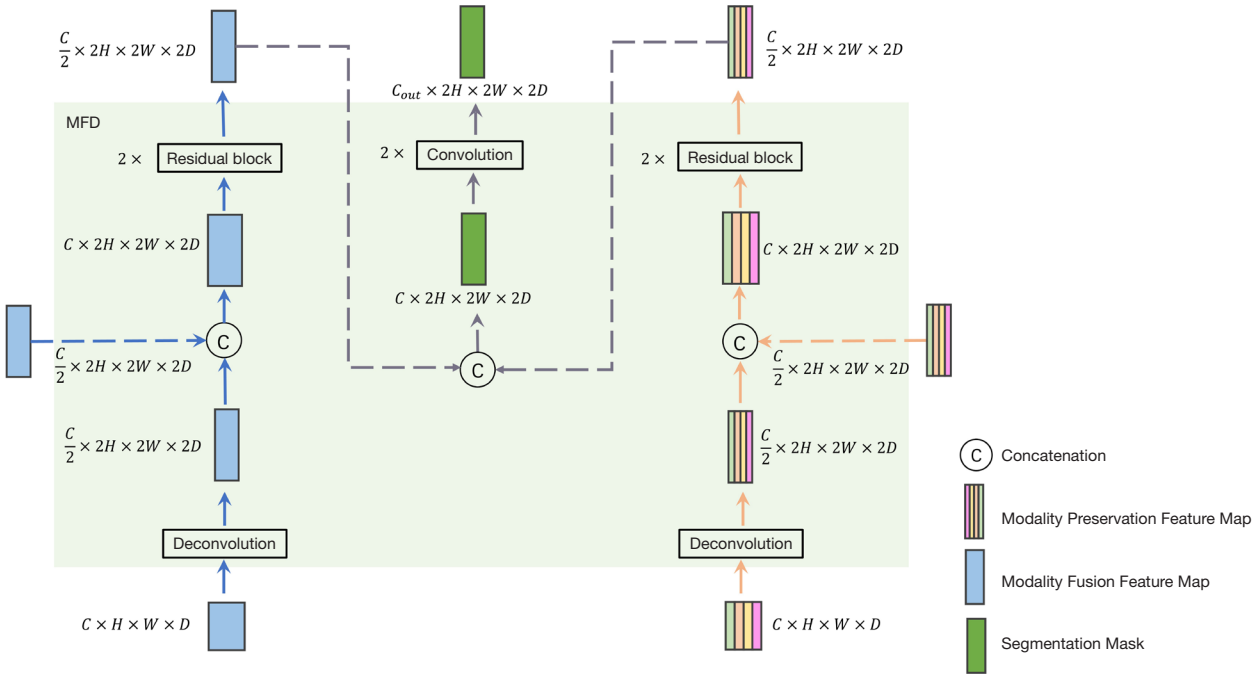

Figure 5 The structure of the MFD. C: channel; H: height; W: width; D: depth. MFD, modality fusion decoder.

in the multimodal fusion channel, and thus obtain better feature representation.

For both single-modal and multimodal feature analysis, the basic component of MPE was the residual block, as shown in *Figure 4B*. The residual block contained two successive convolutions, Batch Normalization and LeakyReLU, and the original feature map was added with the manipulated feature map as the final output of the residual block. For the single-modal part, the convolution in the residual block used group convolution to ensure that the features were not mixed among the modalities. For the multimodal part, the convolution in the residual block adopted common convolution so that the features could be fully mixed. Because the input medical images were in three dimensions, the convolution was all in three dimensions. The size of the convolution kernel was 3×3×3. For the single-modal feature map, MPE contained two contiguous residual blocks (38). For the multimodal feature map, MPE first concatenated the single-modal feature map and the multimodal feature map in the channel dimension and then passed through two consecutive residual blocks for feature extraction. The input single-modal and multimodal feature maps were both C×H×W×D in size (C: channel; H: height; W: width; D: depth), and after MPE, the feature maps became 2C×H/2×W/2×D/2.

MFD

MPU-Net's decoder consisted of four MFD layers. The MFD had a total of four input feature maps: the modality preservation feature map and the modality fusion feature map output by the previous MFD or the last MPE, the modality preservation feature map and modality fusion feature map output by the skip-connected MPE or the first group convolution and common convolution, as shown in *Figure 5*. The three output feature maps of the MFD were a modality preservation feature map, a modality fusion feature map, and a segmentation mask. The modality preservation feature map and modality fusion feature map were used as the input of the next MFD layer, and the segmentation feature map was used for deep supervision training and as the output result. Within the MFD, the input modality preservation feature map and modality fusion feature map were first deconvolved to halve the number of channels but double the spatial dimension, and then concatenated with the corresponding feature map from the skip-connection of the previous MPE, respectively. Next, the concatenated feature map was immediately run through the residual block twice. The structure of the residual block was the same as that of the residual block in the MPE module, but the difference is that the size of the output map remains the same and the channel is reduced by half, as shown in

Figure 5. The modality preservation feature map and the modality fusion feature map obtained at this point could be used as input for the next MFD layer. The two feature maps were concatenated together and output as the segmentation mask of this MFD after two convolutions. During the MFD calculation, it preserved the independence between individual modalities for the modality preservation feature map. For the output segmentation mask, MFD combined the contents of the modality preservation feature map and modality fusion feature map.

Dataset

The multimodal brain tumor segmentation dataset of MSD is a publicly available dataset for brain glioma segmentation from BraTS challenge (25,39). The dataset aimed at evaluating state-of-the-art methods for the segmentation of brain tumors and contained 484 magnetic resonance (MR) cases with segmented labels. The MR images were manually annotated by both clinicians and board-certified radiologists. For each case, T1-weighted (T1), T1-weighted contrast-enhanced (T1ce), T2-weighted (T2) and fluid attenuated inversion recovery (FLAIR) images were provided. The spatial dimensions of each case were 4×155×240×240. In the encoder, the stride of the first convolution of each MPE module was 2 in all three dimensions, while the stride of all other convolutions was 1. In the decoder, the stride of deconvolution in each MFP was 2, and the stride of convolution was 1. The goal of brain tumor segmentation was to segment three different regions, namely, the whole tumor (WT), tumor core (TC) and enhancing tumor (ET). An example diagram of brain tumor image is shown in *Figure 1A*.

Datasets for prostate segmentation of MSD were provided by Radboud University Medical Center (25). The dataset contained 48 MR cases with segmentation labels, each including both the T2 and the apparent diffusion coefficient (ADC) modalities. The spatial dimensions of each case were 2×20×320×320. The segmentation targets of the prostate dataset included the peripheral zone (PZ) and transitional zone (TZ). An example diagram of prostate is shown in *Figure 1B*. Following the UNETR's method (26), the brain tumor and prostate datasets were divided into training, validation and test sets. A random 10% of the data was used as the test set and the rest of the data was used to train the model using five-fold cross-validation. We have more test samples and fewer training samples than UNETR (UNETR's test set represents 5% of all data). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Training

We did not use any existing backbone in MPU-Net, so we trained MPU-Net from scratch. All the convolution layers of MPU-Net were initialized by Glorot (40). In the image preprocessing stage, we performed data augmentation of rotation, scale, and Gaussian noise on brain tumor and prostate data. Specifically, for brain tumor data, three labels were merged into three regions to better segment the tumor (22,41). The whole tumor consists of WT, TC, and ET, the tumor core consists of TC and ET, and the enhancing tumor consists of ET.

During model training, we used SGD Optimizer (42) (learning rate of 0.001, weight_decay =3e-05, momentum =0.99), batch size of 2, and epoch of 200.

Deep supervision was also utilized to solve the problem of information loss during forward propagation and to improve detail accuracy. Auxiliary loss is a technique used in deep learning training to improve the performance of the model. It is an additional loss function that is used alongside the main loss function to help guide the training process. Additional auxiliary losses were added in the decoder to all but the two lowest resolutions (24,43). The weights of the auxiliary losses were decreased by a factor of 2 and normalized for all layers:

$$L = 0.57L_1 + 0.29L_2 + 0.14L_3 \quad [1]$$

The loss function (L) consists of three parts. The first part (L_1) is the difference between the segmentation mask output from the fourth MFDs of MPU-Net and ground truth, the second part (L_2) is the difference between the segmentation mask output from the third MFDs of MPU-Net and ground truth with the same resolution, and the third part (L_3) is the difference between the segmentation mask output from the second MFDs of MPU-Net and ground truth with the same resolution. Each auxiliary loss was composed of a Cross Entropy (CE) Loss (44) and a Dice Loss (45):

$$L_i = CE_i + Dice_i \quad [2]$$

The calculation of *CE* and Dice included the results predicted by the model (p) and the ground truth (gt). Assuming a predicted image with M voxels, the *CE* was

$$CE = -\sum_{k=1}^{M}\left(gt_k \times \log p_k\right)\sum_{k=1}^{M}\left[\left(1-gt_k\right)\times\log\left(1-p_k\right)\right] \quad [3]$$

and the *Dice* was

Table 2 Brain tumor segmentation results of the validation dataset

Method	Dice				HD95			
	WT	TC	ET	Mean	WT	TC	ET	Mean
MPU-Net	0.9092	0.8665	0.8552	0.8770	3.449	4.708	3.579	3.912

Higher Dice scores indicate better results, while lower HD95 scores indicate better results. WT, whole tumor; TC, tumor core; ET, enhancing tumor; HD95, 95% Hausdorff distance; MPU-Net, Modality Preserving U-Net.

Table 3 Brain tumor segmentation results of the test dataset

Method	Dice				HD95			
	WT	TC	ET	Mean	WT	TC	ET	Mean
MPU-Net	0.8942	0.8692	0.8459	0.8698	3.530	4.899	2.555	3.661

Higher Dice scores indicate better results, while lower HD95 scores indicate better results. WT, whole tumor; TC, tumor core; ET, enhancing tumor; HD95, 95% Hausdorff distance; MPU-Net, Modality Preserving U-Net.

$$Dice = \frac{2\times\sum_{k=1}^{M}\left(p_k \times gt_k\right)}{\sum_{k=1}^{M}\left(p_k^2 + gt_k^2\right)} \quad [4]$$

The proposed method was implemented in PyTorch 1.10 and was trained with an AMD EPYC-7302 1.5G × 64 CPU and a NVIDIA RTX A6000 GPU.

Evaluation criteria

To test the performance of our proposed method, we employed two quantitative metrics: Dice score (DSC) and 95% Hausdorff distance (HD95) (46). DSC measures the amount of overlap between the ground truth segmentation and the automated segmentation. Although the HD typically evaluates the difference between two different representations of 3D objects, the HD95 is used in practice rather than HD, as it is harsh and sensitive to noise. DSC was defined in Eq. [4], and *HD* could be calculated as follows:

$$HD = \max\left[\max_{x\in P}\min_{y\in GT} d(x,y), \max_{y\in GT}\min_{x\in P} d(x,y)\right] \quad [5]$$

HD95 and HD were calculated in the same way. However, HD95 was based on the calculation of the 95th percentile of the distances between boundary points in x and y, which means HD95 could eliminate the impact of a very small subset of the outliers (47).

Ablation study

To validate the effectiveness of MPU-Net, we conducted an ablation study on the brain tumor and prostate datasets. In the ablation study, we investigated the effects of different configurations of single-modal preservation and multimodal fusion strategies. For fairness of comparison, when using single-modal or multimodal alone, the number of feature channels in each layer was doubled to ensure the same number of features as when single-model and multimodal were combined. If only the single-modal preservation method was used, the multimodal fusion channel in MPE and MFD was disabled, and vice versa.

Results

Brain tumor segmentation

The MPU-Net was first evaluated by the brain tumor dataset (39). The segmentation results of the validation and test datasets are shown in *Tables 2*,*3*, respectively. We calculated the DSC and HD95, where a larger DSC or smaller HD95 implies a greater similarity between the predicted image and the ground truth. In the validation dataset, the DSCs of MPU-Net for WT, TC and ET were 90.92%, 86.65%, and 85.52%, respectively. The HD95 values of MPU-Net for WT, TC and ET were 3.449, 4.708 and 3.579, respectively. In the test dataset, the DSCs of MPU-Net for WT, TC and ET were 89.42%, 86.92%, and 84.59%, respectively. The HD95 values of MPU-Net for WT, TC and ET were 3.530, 4.899 and 2.555, respectively. We also compared our method with the state-of-the-art method, and the experimental results are shown in *Table 4*. For our method, the average DSC for the three regions was

Table 4 Comparison results of the proposed approach and the other advanced methods on the brain tumor segmentation dataset

Methods	Dice				HD95			
	WT	TC	ET	Mean	WT	TC	ET	Mean
UNETR (26)	0.789	0.761	0.585	0.711	8.266	8.845	9.354	8.822
Novel CNN architecture (48)	0.88	0.79	0.73	0.85	–	–	–	–
Encoder-Decoder-VAE (22)	0.8839	0.8154	0.7664	0.8219	5.9044	4.8091	3.7731	4.8262
Context aware deep learning (49)	0.895	0.835	0.821	0.850	4.897	6.712	3.319	4.976
Two-stage cascaded U-Net (50)	0.8880	0.8370	0.8327	0.8526	4.618	4.131	2.651	3.801
nnU-Net (41)	0.9118	0.8571	0.7985	0.8558	3.73	5.64	26.41	11.93
Ours	0.8942	0.8692	0.8459	0.8698	3.530	4.899	2.555	3.661

Higher Dice scores indicate better results, while lower HD95 scores indicate better results. UNETR, UNEt TRansformers; CNN, convolutional neural network; WT, whole tumor; TC, tumor core; ET, enhancing tumor; HD95, 95% Hausdorff distance; VAE, Variational Autoencoder.

86.98%, which was improved by 1.4% over the previous advanced methods. Compared to other advanced works, MPU-Net showed a greater advantage in both metrics and a significant improvement in segmentation results.

For the qualitative analysis, we showed the segmentation results of MPU-Net for different sites of brain tumors. *Figure 6* shows a visual comparison of brain tumor segmentation in transverse, coronal and sagittal views for different cases. It was clear from *Figure 6* that MPU-Net could generate highly accurate segmentation masks for different sites, volumes and morphologies of brain tumors.

Prostate segmentation

The MPU-Net was also tested using the prostate dataset (25). The segmentation results of the validation and test datasets are shown in *Tables 5*,*6*, respectively. In the validation dataset, the average DSCs of MPU-Net for PZ and TZ were 72.60% and 90.11%, and the HD95 for PZ and TZ were 5.807 and 4.925, respectively. In the test dataset, the average DSCs of MPU-Net for PZ and TZ were 71.20% and 90.38%, and the HD95 for PZ and TZ were 6.367 and 4.766, respectively. We also compared our method with the state-of-the-art method, and the experimental results are shown in *Table 7*. Compared with other advanced methods, MPU-Net had the best result in the TZ segmentation and ranked fourth in the average. The segmentation results of MPU-Net for the prostate are shown in *Figure 7*.

Ablation study

The results of the ablation study are shown in *Tables 8*,*9*. When using only the single-modal preservation method, there was no fusion between the individual image modalities, i.e., the independence of the individual modalities was strictly preserved. The results of brain tumor segmentation are shown in the first row of *Table 8*, where the DSCs of WT, TC and ET were 0.8865, 0.8214, and 0.8111, respectively, and the mean DSC was 0.8396. The results of prostate segmentation are shown in the first row of *Table 9*, where the DSCs of PZ and TZ were 0.7098 and 0.8727, respectively, and the mean DSC was 0.7913. In the case where only multimodal fusion was used, the individual modalities were all fused, and no independent modality was retained. The results of brain tumor segmentation are shown in the second row of *Table 8*, where the DSCs of WT, TC and ET were 0.8830, 0.8321, and 0.8196, respectively, and the mean DSC was 0.8449. The results of prostate segmentation are shown in the second row of *Table 9*, where the DSCs of PZ and TZ were 0.7113 and 0.8735, respectively, and the mean DSC was 0.7924. If both single-modal preservation and multimodal fusion were used but there was no connection between single-modal preservation and multimodal fusion in the MPE, the results of brain tumor segmentation are shown in the third row of *Table 8*, where the DSCs of WT, TC and ET were 0.9028, 0.8588, and 0.8517, respectively, and the mean DSC was 0.8711. The results of prostate segmentation are shown in the third row of *Table 9*, where the DSCs of PZ and TZ were 0.7211 and 0.8892, respectively, and the mean DSC was 0.8051. If both single-modal preservation and multimodal fusion were used and the connection between single-modal preservation and multimodal fusion was preserved, i.e., the original MPU-Net design, the results of brain tumor segmentation

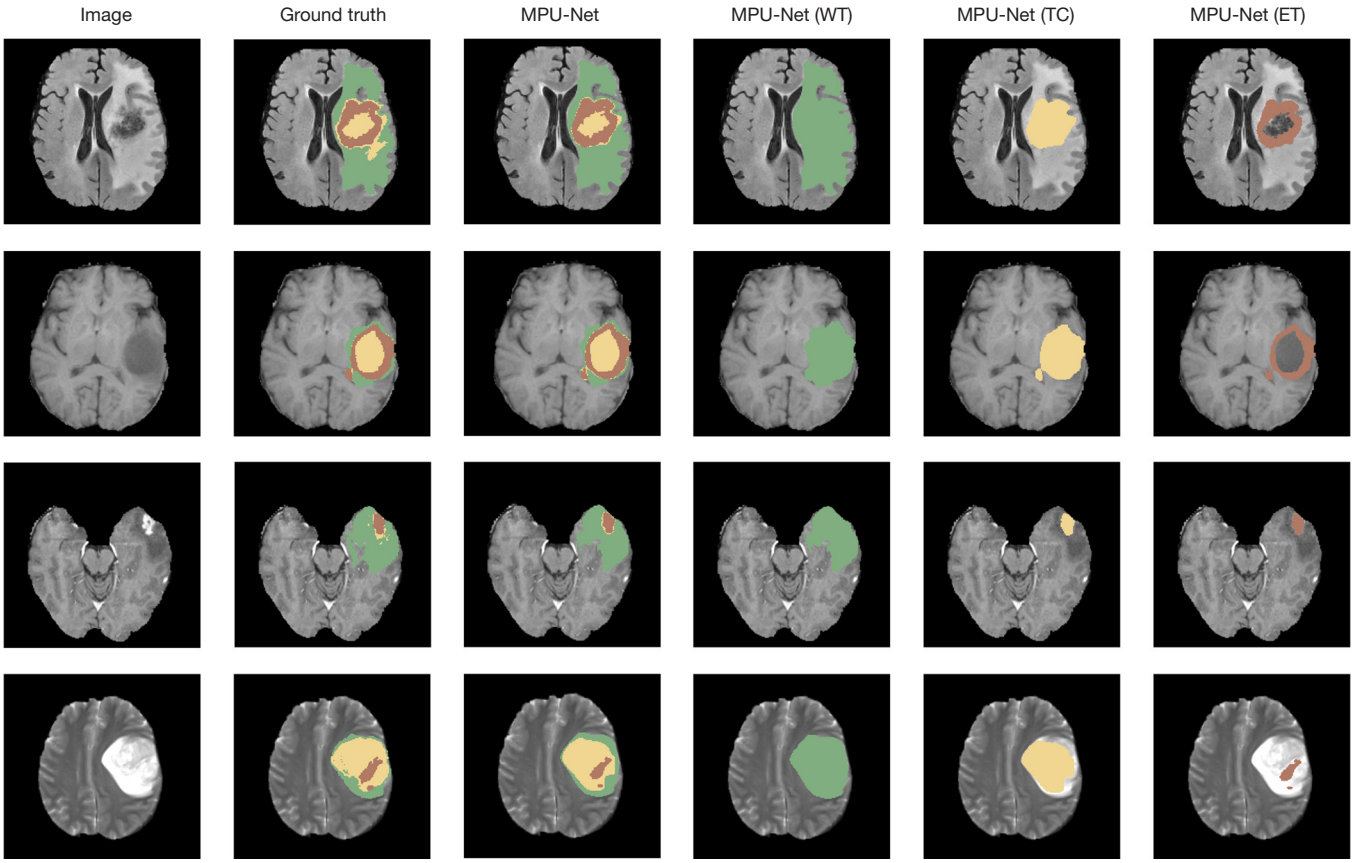

Figure 6 Segmentation results of brain tumor. The first column is the brain tumor image, and the second column is the ground truth of the tumor. The third column is the segmentation result of the MPU-Net. The fourth, fifth and sixth columns are the whole tumor, tumor core, and enhancing tumor from the MPU-Net segmentation, respectively. The four rows of the image are four different cases, and each row shows only one of the four modalities—FLAIR, T1, T1ce and T2 in the first to fourth rows, respectively. All four modalities are used in the actual segmentation process. MPU-Net, Modality Preserving U-Net; WT, whole tumor; TC, tumor core; ET, enhancing tumor; FLAIR, fluid attenuated inversion recovery; T1, T1-weighted; T1ce, T1-weighted contrast-enhanced; T2, T2-weighted.

Table 5 Prostate segmentation results of the validation dataset

Method	Dice			HD95		
	PZ	TZ	Mean	PZ	TZ	Mean
MPU-Net	0.7260	0.9011	0.8136	5.807	4.925	5.366

Higher Dice scores indicate better results, while lower HD95 scores indicate better results. PZ, peripheral zone; TZ, transitional zone; HD95, 95% Hausdorff distance; MPU-Net, Modality Preserving U-Net.

Table 6 Prostate segmentation results of the test dataset

Method	Dice			HD95		
	PZ	TZ	Mean	PZ	TZ	Mean
MPU-Net	0.7120	0.9038	0.8079	6.367	4.766	5.567

Higher Dice scores indicate better results, while lower HD95 scores indicate better results. PZ, peripheral zone; TZ, transitional zone; HD95, 95% Hausdorff distance; MPU-Net, Modality Preserving U-Net.

Table 7 Comparison results of the proposed approach and the other advanced methods on the prostate dataset

Methods	Dice			HD95		
	PZ	TZ	Mean	PZ	TZ	Mean
CNN with a novel feature pyramid attention (51)	0.74	0.86	0.80	–	–	–
Multiple CNN (52)	–	0.847	–	–	–	–
nnU-Net (24)	0.77	0.90	0.835	–	–	–
SwinUNETR (27)	0.82	0.89	0.855	–	–	–
Universal Modal (28)	0.83	0.90	0.865	–	–	–
Ours	0.7120	0.9038	0.8079	6.367	4.766	5.567

Higher Dice scores indicate better results, while lower HD95 scores indicate better results. CNN, convolutional neural network; UNETR, UNEt TRansformers; PZ, peripheral zone; TZ, transitional zone; HD95, 95% Hausdorff distance.

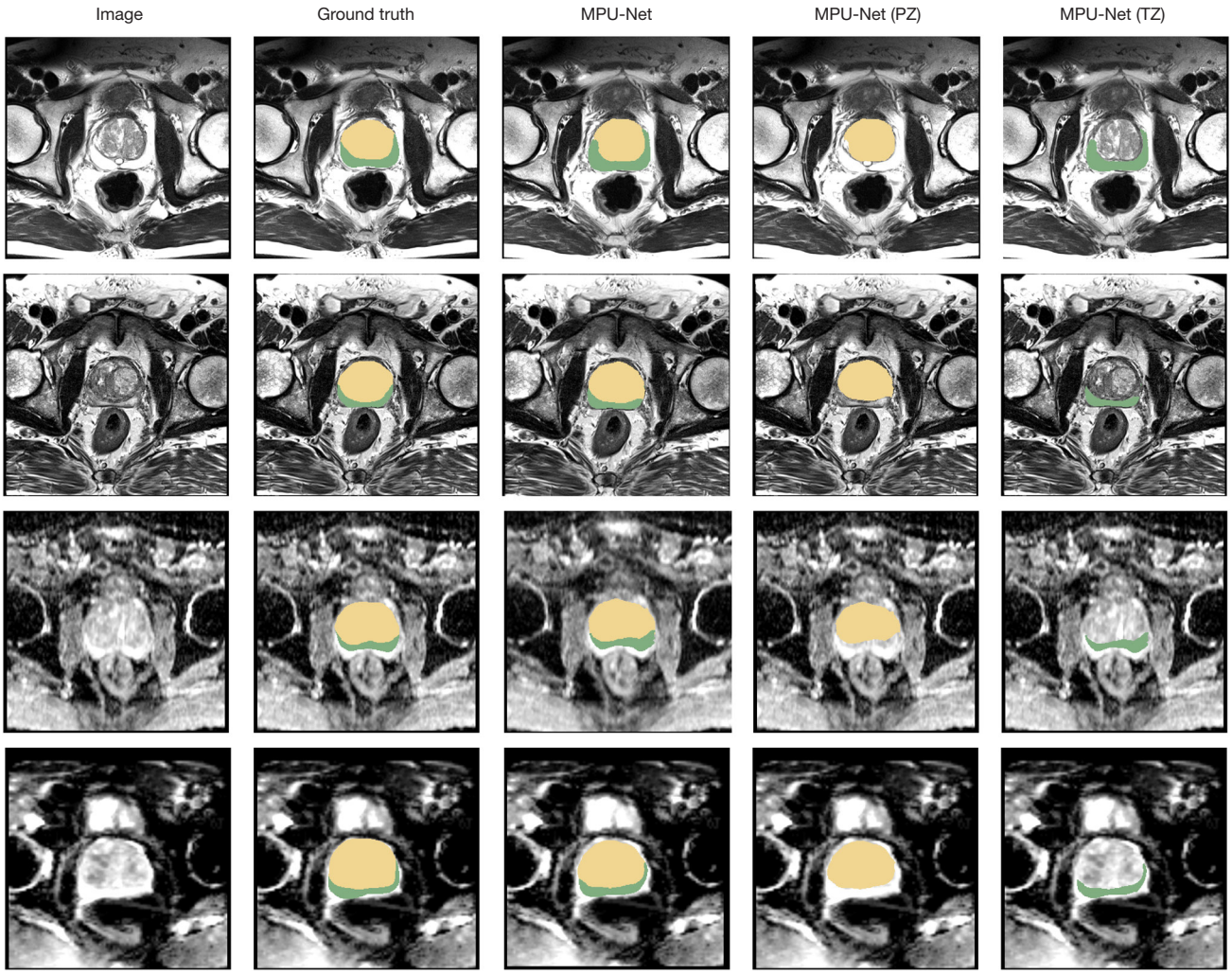

Figure 7 Segmentation results of the prostate. The first column is the prostate image, and the second column is the ground truth of the prostate. The third column is the segmentation result of the MPU-Net. The fourth and fifth columns are the PZ and TZ from the MPU-Net segmentation, respectively. The four rows of the image are four different cases. The first two rows show the MRI modality, and the second two rows show the ADC modality. Both modalities are used in the actual segmentation process. MPU-Net, Modality Preserving U-Net; PZ, peripheral zone; TZ, transitional zone; MRI, magnetic resonance imaging; ADC, apparent diffusion coefficient.

Table 8 Segmentation results of the ablation study (brain tumor dataset)

Methods	Dice			
	WT	TC	ET	Mean
Single-modal preservation	0.8865	0.8214	0.8111	0.8396
Multimodal fusion	0.8830	0.8321	0.8196	0.8449
Single-modal preservation & multimodal fusion	0.9028	0.8588	0.8517	0.8711
MPU-Net	0.9092	0.8665	0.8552	0.8770

WT, whole tumor; TC, tumor core; ET, enhancing tumor; MPU-Net, Modality Preserving U-Net.

Table 9 Segmentation results of the ablation study (prostate dataset)

Methods	Dice		
	PZ	TZ	Mean
Single-modal preservation	0.7098	0.8727	0.7913
Multimodal fusion	0.7113	0.8735	0.7924
Single-modal preservation & multimodal fusion	0.7211	0.8892	0.8051
MPU-Net	0.7260	0.9011	0.8136

PZ, peripheral zone; TZ, transitional zone; MPU-Net, Modality Preserving U-Net.

are shown in the fourth row of *Table 8*, where the DSCs of WT, TC and ET were 0.9092, 0.8665, and 0.8552, respectively, and the mean DSC was 0.8770. The results of prostate segmentation are shown in the fourth row of *Table 9*, where the DSCs of PZ and TZ were 0.7260 and 0.9011, respectively, and the mean DSC was 0.8136. Therefore, the combination of single-modal preservation and multimodal fusion methods significantly improved the effectiveness of multimodal medical image feature analysis.

The analysis of variance (ANOVA) showed that there were significant differences between the four groups ($P<0.001$ in both brain tumor and prostate segmentation). The results of *t*-test showed that the segmentation results of MPU-Net were different from single-modal preservation, multimodal fusion and single-modal preservation & multimodal fusion ($P<0.001$, $P<0.001$ and $P=0.051$ in brain tumor segmentation; $P<0.001$, $P=0.001$ and $P=0.034$ in prostate segmentation). In the *t*-test, the architecture of single-modal preservation & multimodal fusion and MPU-Net do not show very strong differences, because both use the multiscale single-modal preservation and multimodal fusion image analysis strategy proposed in this paper, but differ only in the absence of the connection from single-modal features to multimodal features in the MPE of the single-modal preservation & multimodal fusion architecture.

Discussion

We have introduced MPU-Net, which can combine single-modal and multimodal image information at different scales to achieve the segmentation of multimodal medical images. We assume that preserving the independence of a single modality in the segmentation process and combining a single modality with multiple modalities facilitates the optimization of complex medical image segmentation tasks. The proposed method has been tested on the brain tumor and prostate segmentation dataset and achieved good results. In brain tumor segmentation tasks, the ET region is often difficult to segment, and the appearance of this region is strongly dependent on the T1ce modality. Compared to other advanced methods, MPU-Net has the best result in segmentation for ET. In summary, MPU-Net achieves good results in multimodal medical image segmentation challenges, with enhancements for multiple objectives of medical image segmentation. However, MPU-Net still needs further validation on more datasets, and the miniaturization and rapidity of multimodal medical image segmentation networks should be further investigation. Downstream applications of MPU-Net segmentation, such as using the segmentation results for disease diagnosis or intraoperative navigation, also deserve further exploration.

We proposed a novel multimodal medical image semantic segmentation network, MPU-Net, which can segment medical targets from multimodal medical image data. The network consisted of two main modules, the MPE module and the MFD module, and the feature analysis strategy of single-modal preservation and multimodal fusion was implemented. In the multimodal brain tumor segmentation task, the mean DSC and the mean HD95 of MPU-Net were 86.98% and 3.661, respectively. In the multimodal prostate segmentation task, the mean DSC and the mean HD95 of MPU-Net were 80.79% and 5.567, respectively. In brain tumor segmentation, the proposed MPU-Net achieved the best results in TC and ET segmentation compared with other advanced methods, and ranked first in mean Dice and mean HD95 for all three region segmentations. In prostate segmentation, the proposed MPU-Net achieved

the best results in TZ segmentation compared with other advanced methods, and ranked fourth in mean Dice for two region segmentation. MPU-Net can effectively improve the performance of multimodal medical image segmentation. In the ablation study, the combination of single-modal preservation and multimodal fusion achieved the best results, demonstrating the effectiveness of the proposed MPU-Net.

Conclusions

In this paper, we used two multimodal medical image datasets, brain tumor and prostate segmentation in MSD, and also different versions of brain tumor datasets exist in our experimental results comparison. Since the online test of the MSD challenge requires segmentation of all datasets (including 8 single modal datasets and 2 multimodal datasets), and the proposed method in this paper focuses on multimodal medical image segmentation, it cannot participate in the online test of MSD challenge. We followed UNETR's method (26) and used the local test dataset. On the other hand, two different segmentation regions exist for brain tumor segmentation results, one for WT, TC and ET and the other for edema, ET and necrosis regions (41). In this paper, WT, TC and ET regions are used as segmentation results, so segmentation results of brain tumor can only be compared with the work that uses the same segmentation regions. In future work, as the online validation and evaluation metrics of multimodal medical image datasets continue to be improved and unified, the proposed method should be further investigated and compared with more comprehensive and sophisticated multimodal medical image datasets.

Acknowledgments

We gratefully thank the reviewers for their constructive comments. We gratefully thank the Medical Segmentation Decathlon dataset and the PyTorch, MONAI and nnU-Net architectures.
Funding: This work was supported by the National Key R&D Program of China (No. 2022YFA1104800).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-22-1367/coif). The 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).

References

1. Meola A, Cutolo F, Carbone M, Cagnazzo F, Ferrari M, Ferrari V. Augmented reality in neurosurgery: a systematic review. Neurosurg Rev 2017;40:537-48.
2. Wu B, Liu P, Xiong C, Li C, Zhang F, Shen S, Shao P, Yao P, Niu C, Xu R. Stereotactic co-axial projection imaging for augmented reality neuronavigation: a proof-of-concept study. Quant Imaging Med Surg 2022;12:3792-802.
3. Liu P, Li C, Xiao C, Zhang Z, Ma J, Gao J, Shao P, Valerio I, Pawlik TM, Ding C, Yilmaz A, Xu R. A Wearable Augmented Reality Navigation System for Surgical Telementoring Based on Microsoft HoloLens. Ann Biomed Eng 2021;49:287-98.
4. Wang S, Yang DM, Rong R, Zhan X, Xiao G. Pathology Image Analysis Using Segmentation Deep Learning Algorithms. Am J Pathol 2019;189:1686-98.
5. Thakur N, Juneja M. Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma. Biomedical Signal Processing and Control 2018;42:162-89.
6. Seebock P, Orlando JI, Schlegl T, Waldstein SM, Bogunovic H, Klimscha S, Langs G, Schmidt-Erfurth U. Exploiting Epistemic Uncertainty of Anatomy Segmentation for Anomaly Detection in Retinal OCT. IEEE Trans Med Imaging 2020;39:87-98.
7. Yamanakkanavar N, Choi JY, Lee B. MRI Segmentation and Classification of Human Brain Using Deep Learning for Diagnosis of Alzheimer's Disease: A Survey. Sensors (Basel) 2020;20:3243.
8. Middlebrooks EH, Tuna IS, Grewal SS, Almeida L,

Heckman MG, Lesser ER, Foote KD, Okun MS, Holanda VM. Segmentation of the Globus Pallidus Internus Using Probabilistic Diffusion Tractography for Deep Brain Stimulation Targeting in Parkinson Disease. AJNR Am J Neuroradiol 2018;39:1127-34.
9. Chen X, Williams BM, Vallabhaneni SR, Czanner G, Williams R, Zheng Y. Learning active contour models for medical image segmentation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019:11632-40.
10. Zhang Y, Wu J, Liu Y, Chen Y, Chen W, Wu EX, Li C, Tang X. A deep learning framework for pancreas segmentation with multi-atlas registration and 3D level-set. Med Image Anal 2021;68:101884.
11. Jiang Y, Zhao K, Xia K, Xue J, Zhou L, Ding Y, Qian P. A Novel Distributed Multitask Fuzzy Clustering Algorithm for Automatic MR Brain Image Segmentation. J Med Syst 2019;43:118.
12. Ouyang C, Biffi C, Chen C, Kart T, Qiu H, Rueckert D. Self-supervision with Superpixels: Training Few-Shot Medical Image Segmentation Without Annotation. In: Vedaldi A, Bischof H, Brox T, Frahm JM. editors. Computer Vision – European Conference on Computer Vision 2020. Cham: Springer, 2020:762-80.
13. Chen X, Pan L. A Survey of Graph Cuts/Graph Search Based Medical Image Segmentation. IEEE Rev Biomed Eng 2018;11:112-24.
14. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016:770-8.
15. Zhou T, Ruan S, Canu S. A review: Deep learning for medical image segmentation using multi-modality fusion. Array 2019;3-4:100004.
16. Guo Z, Li X, Huang H, Guo N, Li Q. Deep Learning-based Image Segmentation on Multimodal Medical Imaging. IEEE Trans Radiat Plasma Med Sci 2019;3:162-9.
17. Yang Y, Wu J, Huang S, Fang Y, Lin P, Que Y. Multimodal Medical Image Fusion Based on Fuzzy Discrimination With Structural Patch Decomposition. IEEE J Biomed Health Inform 2019;23:1647-60.
18. Lee SY, Jeon SI, Jung S, Chung IJ, Ahn CH. Targeted multimodal imaging modalities. Adv Drug Deliv Rev 2014;76:60-78.
19. 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. Cham: Springer, 2015:234-41.
20. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, et al. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. arXiv 2019. arXiv:1811.02629.
21. Kamnitsas K, Bai W, Ferrante E, McDonagh S, Sinclair M, Pawlowski N, Rajchl M, Lee M, Kainz B, Rueckert D, Glocker B. Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation. In: Crimi A, Bakas S, Kuijf H, Menze B, Reyes M. editors. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Cham: Springer, 2017.
22. Myronenko A. 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T. editors. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Cham: Springer, 2019:311-20.
23. Qin X, Zhang Z, Huang C, Dehghan M, Zaiane OR, Jagersand M. U2-Net: Going deeper with nested U-structure for salient object detection. Pattern Recognition 2020;106:107404.
24. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203-11.
25. Antonelli M, Reinke A, Bakas S, Farahani K, Kopp-Schneider A, Landman BA, et al. The Medical Segmentation Decathlon. Nat Commun 2022;13:4128.
26. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, Roth H, Xu D. UNETR: Transformers for 3D Medical Image Segmentation. 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2022.
27. Tang Y, Yang D, Li W, Roth H, Landman B, Xu D, Nath V, Hatamizadeh A. Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. arXiv 2022. arXiv:2111.14791.
28. Liu J, Zhang Y, Chen JN, Xiao J, Lu Y, Landman BA, Yuan Y, Yuille A, Tang Y, Zhou Z. CLIP-Driven Universal Model for Organ Segmentation and Tumor Detection arXiv 2023. arXiv:2301.00785.
29. Srivastava N, Salakhutdinov R. Multimodal Learning with Deep Boltzmann Machines. Journal of Machine Learning Research 2014;15:2949-80.
30. Yang J, Beadle BM, Garden AS, Schwartz DL, Aristophanous M. A multimodality segmentation

framework for automatic target delineation in head and neck radiotherapy. Med Phys 2015;42:5310-20.
31. Zhang W, Li R, Deng H, Wang L, Lin W, Ji S, Shen D. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. Neuroimage 2015;108:214-24.
32. Dolz J, Gopinath K, Yuan J, Lombaert H, Desrosiers C, Ben Ayed I. HyperDense-Net: A Hyper-Densely Connected CNN for Multi-Modal Image Segmentation. IEEE Trans Med Imaging 2019;38:1116-26.
33. Nie D, Wang L, Gao Y, Shen D. Fully convolutional networks for multi-modality isointense infant brain image segmentation. Proc IEEE Int Symp Biomed Imaging 2016;2016:1342-5.
34. Huang B, Yang F, Yin M, Mo X, Zhong C. A Review of Multimodal Medical Image Fusion Techniques. Comput Math Methods Med 2020;2020:8279342.
35. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. arXiv 2021. arXiv:2105.05537.
36. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. IEEE Trans Med Imaging 2020;39:1856-67.
37. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM 2017;60:84-90.
38. Zagoruyko S, Komodakis N. Wide Residual Networks. arXiv 2016. arXiv:1605.07146.
39. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data 2017;4:170117.
40. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. International Conference on Artificial Intelligence and Statistics 2010.
41. Isensee F, Jaeger PF, Full PM, Vollmuth P, Maier-Hein KH. nnU-Net for Brain Tumor Segmentation. arXiv 2020. arXiv:2011.00848.
42. Amari S. Backpropagation and stochastic gradient descent method. Neurocomputing 1993;5(4):185-96.
43. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 201:1-9.
44. Cox DR. The regression analysis of binary sequences. Journal of the Royal Statistical Society: Series B (Methodological) 1958;20:215-32.
45. Soomro TA, Afifi AJ, Gao J, Hellwich O, Paul M, Zheng L. Strided U-Net Model: Retinal Vessels Segmentation using Dice Loss. 2018 Digital Image Computing: Techniques and Applications (DICTA). Canberra: IEEE, 2018:1-8.
46. Jesorsky O, Kirchberg KJ, Frischholz RW. Robust face detection using the hausdorff distance. In: Bigun J, Smeraldi F. editors. Audio- and Video-Based Biometric Person Authentication. AVBPA 2001. Berlin, Heidelberg: Springer-Verlag, 2001:90-5.
47. Ghaffari M, Sowmya A, Oliver R. Automated Brain Tumour Segmentation Using Cascaded 3D Densely-Connected U-Net. In: Crimi A, Bakas S. editors. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2020. Cham: Springer, 2021:481-91.
48. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H. Brain tumor segmentation with Deep Neural Networks. Med Image Anal 2017;35:18-31.
49. Pei L, Vidyaratne L, Rahman MM, Iftekharuddin KM. Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images. Sci Rep 2020;10:19726.
50. Jiang Z, Ding C, Liu M, Tao D. Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task. In: Crimi A, Bakas S. editors. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2019. Cham: Springer, 2020:231-41.
51. Liu Y, Yang G, Mirak SA, Hosseiny M, Azadikhah A, Zhong X, Reiter RE, Lee Y, Raman SS, Sung K. Automatic Prostate Zonal Segmentation Using Fully Convolutional Network with Feature Pyramid Attention. IEEE Access 2019;7:163626-32.
52. Clark T, Zhang J, Baig S, Wong A, Haider MA, Khalvati F. Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks. J Med Imaging (Bellingham) 2017;4:041307.

Cite this article as: Wu B, Zhang F, Xu L, Shen S, Shao P, Sun M, Liu P, Yao P, Xu RX. Modality preserving U-Net for segmentation of multimodal medical images. Quant Imaging Med Surg 2023;13(8):5242-5257. doi: 10.21037/qims-22-1367

