MCA-UNet: A Multi-Center Adaptive U-Net for Improved Tumor Segmentation in Multi-Center PET/CT Images

1st Meng Wang
School of Electronic and Information Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
wangmeng_dx@lntu.edu.cn

2nd Yongwei Tang
School of Electronic and Information Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
2458672708@qq.com

3rd Yarong Feng
School of Electronic and Information Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
18093490172@163.com

4th Yuxin Liang
School of Electronic and Information Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
2833391387@qq.com

5th Tianyu Shi*
School of information Science and Engineering
Shenyang Ligong University
Shenyang city, 110159, P.R.China
ty.shi@sylu.edu.cn

6th Chao Lv*
Department of General Surgery, Shengjing Hospital
China Medical University
Shenyang city, 110159, P.R.China
clu@cmu.edu.cn

***Abstract*—The performance of medical image segmentation models faces the core challenge of data heterogeneity in multi-center collaborative training. Existing methods typically process mixed multi-center data directly while ignoring center-specific characteristics, resulting in suboptimal performance. To address this issue, this study proposes a Multi-Center Adaptive U-Net (MCA-UNet) for PET/CT images. The proposed method introduces a hierarchical center adaptation mechanism: center-specific adapters are employed in shallow layers to handle low-level variations, while multi-branch cross-center blocks in deeper layers balance general features and center-specific representations. In addition, MCA-UNet incorporates a low-rank dual-path fusion strategy to further enhance overall performance. Extensive experiments were conducted on the publicly available HECKTOR 2021 head and neck tumor segmentation dataset. Compared with other methods, MCA-UNet achieves superior performance, with a Dice score of 81.50% and an HD95 of 5.48 mm, with limited parameter growth (less than 5%). These results demonstrate that the proposed method can significantly improve overall segmentation performance on multi-center PET/CT images.**

Keywords—Tumor Segmentation, Multi-Center Data, PET/CT, Low-Rank Matrix, U-Net

I. INTRODUCTION

Tumor segmentation based on positron emission tomography/computed tomography (PET/CT) is a common yet challenging task in medical imaging. Computed tomography (CT) provides high-resolution anatomical structures but still struggles to distinguish tumors from surrounding tissues [1]; positron emission tomography (PET) reflects tumor metabolic activity, thereby enhancing tumor identification [2], but suffers from relatively low resolution. Therefore, integrating PET and CT images to fuse anatomical and metabolic information has emerged as an effective approach to improve tumor segmentation accuracy [3]. With the advancement of deep learning techniques, medical image segmentation methods based on convolutional neural networks (especially U-Net [4] and its variants) have achieved significant progress in PET/CT tumor segmentation [5-7]. These methods typically rely on large amounts of training data. Given the scarcity and diversity of medical imaging data, collecting sufficient samples from a single hospital is challenging and insufficient for constructing high-quality datasets [8]. Therefore, there is a need in the clinical community to effectively integrate data from multiple centers to facilitate the development of robust deep learning models. However, applying these advanced models to multi-center PET/CT medical data poses a risk of overall performance degradation. On the one hand, multi-center medical imaging data typically originate from different medical devices, acquisition protocols, reconstruction algorithms, and patient populations, leading to significant variations in data distributions [9]. On the other hand, existing methods usually train a single model directly on multi-center data; however, due to conflicting data distributions across centers, their generalization ability is severely limited [10].

To enhance the overall performance of multi-center PET/CT image segmentation tasks, this study proposes a Multi-Center Adaptive U-Net (MCA-UNet) based on PET/CT images. This model incorporates an adaptive mechanism specifically designed to address variations in multi-center data distributions. It effectively adapts to the data characteristics of different centers without substantially increasing model parameters, thereby improving overall model performance. Specifically, we construct an enhanced encoder-decoder architecture by systematically integrating the Dataset-Aware Block (DAB) [11] with the Multi-Branch Cross-Center Convolutional Block

*Corresponding author.

(MBCCB). At the shallow encoder layers, a Center-Specific Adapter (CSA) captures low-level image detail variations across centers, such as scanning protocols and image reconstruction. At the bottleneck layer, a Dual-Path Feature Fusion Module (DPFFM) achieves adaptive fusion of general features and center-specific features at a higher abstraction level. Compared to the classic U-Net baseline, MCA-UNet effectively balances multi-center data adaptability and parameter efficiency with only a slight increase in parameters, providing a viable solution for multi-center medical image segmentation.

In summary, the main contributions of this paper include:

• We innovatively design a multi-center adaptive U-Net architecture, MCA-UNet. Based on the classic encoder-decoder structure, it incorporates Multi-Branch Cross-Center Convolutional Blocks into the encoder and Dataset-Aware Blocks into the decoder. This approach preserves the general anatomical information across multi-center datasets while adapting to the distribution characteristics of different datasets through center-specific batch normalization layers. Consequently, it enhances the cross-center reusability and representational capability of model features.

• We propose a hierarchical center adaptation mechanism: at shallow layers of the encoder, center-specific adapters are introduced to capture low-level image detail differences across centers; at deeper layers, the center-specific Batch Normalization (BN) layer within the MBCCB module is employed to address cross-center distribution shifts in high-level semantic features.

• We design a dual-path feature fusion module consisting of a shared bottleneck layer and a cross-center low-rank bottleneck layer. The outputs of both branches are adaptively fused via a learnable fusion weight, enabling a parameter-efficient balance between generic representation learning and center-specific modeling.

II. Related Work

A. PET/CT Image Segmentation

Early PET/CT research primarily relied on traditional image processing methods and machine learning techniques, such as threshold segmentation [12] and region growing [13]. While these approaches possess clear physical significance, they often depend on manually designed features, making them difficult to adapt to complex medical imaging scenarios [14]. With the rise of convolutional neural networks (CNNs), deep learning-based medical image segmentation methods gradually became mainstream[5,15-16]. The fully convolutional network (FCN) proposed by Long et al. [17] first achieved end-to-end semantic segmentation, offering new insights for medical image analysis. Building upon this, the U-Net architecture proposed by Ronneberger et al. [4] achieved remarkable success in medical image segmentation tasks through its symmetric encoder-decoder structure and skip connections. In recent years, numerous improvements to U-Net have emerged. Zhou et al. [18] introduced UNet++, which further enhances feature fusion capabilities by incorporating dense skip connections. Çiçek et al. [19] pioneered the extension of U-Net to 3D medical image segmentation by proposing 3D U-Net. These successes have also promoted applications in PET/CT tumor segmentation; however, they are primarily focused on single-center data and show limited performance when handling multi-center datasets.

B. Cross-Center Learning

In medical imaging, even within the same modality, equipment and imaging protocols from different manufacturers can lead to cross-center variability in images. This is similarly true for natural images. Given the high cost of medical imaging, fully leveraging multi-center datasets holds significant clinical value. Wang et al. [11] employed center-generic convolutional layers and center-specific Batch Normalization layers to decouple and learn multi-center data distribution variations alongside generic features, achieving improved overall segmentation performance. However, these approaches have not been applied to medical image segmentation. Domain adaptation methods [20-23] aim to transfer models trained at center A to center B to enhance inference performance on center B data. Yet these approaches cannot simultaneously leverage multi-center data, imposing certain clinical limitations. In contrast, our method integrates multi-center learning with medical image data, thereby improving the performance and efficiency of deep learning on multi-center medical image datasets.

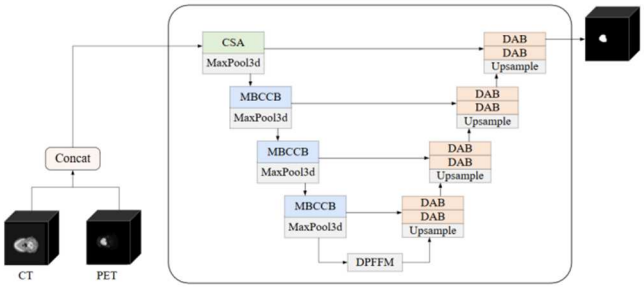

Fig. 1. Overall architecture of MCA-UNet

III. Method

A. Overall Architecture

This paper proposes MCA-UNet, an enhanced U-Net architecture tailored for multi-center medical image segmentation tasks. Fig. 1 illustrates the proposed architecture. The overall structure comprises five components: an input layer, four encoder stages, a bottleneck layer, four decoder stages, and an output layer. The network takes dual-channel PET/CT images as input and outputs tumor segmentation maps.

1) Input Layer: Accepts dual-channel PET and CT images as input, concatenates them along the channel dimension, and feeds the fused features into the encoder.

2) Encoder Stage: The first encoder layer incorporates a CSA module to capture low-level image detail differences across centers. The subsequent three encoder layers utilize MBCCB modules to process inter-domain variations in high-level semantic features across centers.

3) Bottleneck Layer: A dual-path feature fusion module is employed to construct a dual-bottleneck architecture with a cascaded dual-branch design, where the shared path learns cross-center general feature representations, while the low-rank

cross-center path captures center-specific features associated with each center.

4) Decoder Stage: Each decoder layer incorporates a DAB module that progressively restores spatial resolution while adaptively handling the heterogeneity of multi-center data.

B. Dataset-Aware Block

DAB forms the foundation of MCA-UNet's center-specific modeling capability. We employ DAB within the decoder, whose structure is illustrated in Fig. 2. The module employs a 3D convolutional layer for feature extraction. The convolution kernel size is 3×3×3 with a padding setting of 1 to preserve the spatial resolution of feature maps. To address variations in multi-center data distributions, we utilize center-specific batch normalization layers, with each participating center equipped with an independent BN layer. During forward propagation, the system automatically selects the corresponding normalization layer based on the center identifier of the input image. This process can be formally expressed as:

$$y = \mathrm{ReLU}(\mathrm{BN}_c(\mathrm{Conv}(x))) \tag{1}$$

Where x denotes the input feature map, c represents the center identifier, and BN_c indicates the specific batch normalization layer selected based on c.

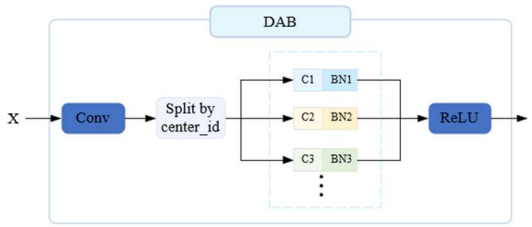

Fig. 2. Dataset-Aware Block

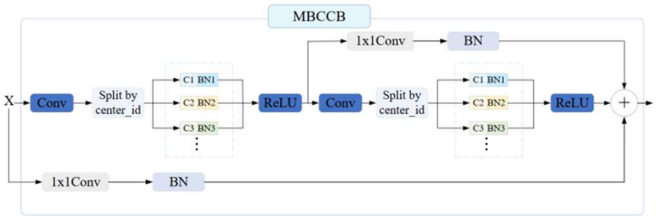

Fig. 3. Multi-Branch Cross-Center Convolutional Block

C. Multi-Branch Cross-Center Convolutional Block

Building upon DAB, we propose the MBCCB module, as shown in Fig. 3. The core of this module consists of two consecutive 3D convolutional layers. Each convolutional layer is followed by center-specific Batch Normalization and ReLU activation to explicitly model the distributional differences of multi-center data. Simultaneously, the module introduces dual residual connections: a long-connection residual path (directly connecting from the module input to the output) and a short-connection residual path (connecting the output of the first convolutional layer to the output of the second). When the number of input and output channels is inconsistent, the residual paths are aligned in channel dimension using 1×1 convolutions to ensure feature additivity.

The dual residual design, while preserving the adaptability of center-specific BN to cross-center data distributions, constructs an efficient pathway for universal feature transfer. The raw input of MBCCB is directly passed to subsequent layers through the long-connection residual path. These features contain general anatomical structural information independent of center, such as the shape and positional relationships of organs. This type of information exhibits high consistency across different datasets and is crucial for segmentation tasks. Long-connection residual paths ensure that this crucial information is not excessively distorted or lost in deep networks, promoting the continuous flow and accumulation of general features. Simultaneously, short-connection residual paths superimpose the output of the first convolutional layer onto the output of the second convolutional layer, effectively mitigating the vanishing gradient problem in deep networks. This design not only stabilizes the training process but also promotes the reuse of intermediate layer features, enabling the model to learn center-specific features while maintaining effective transfer of general features. Through the synergistic effect of long and short residuals, the model can effectively learn the common representations of the dataset while adapting to differences between centers, enhancing the convergence stability and segmentation accuracy of training. Specifically, given input features $x \in R^{C \times D \times H \times W}$, the computational process of the MBCCB module can be represented as follows:

1) Long-Connection Residual Path

$$X = \begin{cases} x, & \text{if channels match} \\ \mathrm{Conv}_{1\times1\times1}(x), & \text{otherwise} \end{cases} \tag{2}$$

2) Primary Path First Level

$$F_1 = \mathrm{ReLU}\big(\mathrm{BN}_c(\mathrm{Conv}_{3\times3\times3}(x))\big) \tag{3}$$

Where c denotes the center identifier, and BN_c represents the specific batch normalization layer selected based on c.

3) Second Level of the Main Path

$$F_2 = \mathrm{ReLU}\big(\mathrm{BN}_c(\mathrm{Conv}_{3\times3\times3}(F_1))\big) \tag{4}$$

4) Shortest Path Residual

$$F_3 = \begin{cases} F_1, & \text{if channels match} \\ \mathrm{Conv}_{1\times1\times1}(F_1), & \text{otherwise} \end{cases} \tag{5}$$

5) Overall Formula

$$Y = F_3 + F_2 + X \tag{6}$$

This work was supported by the National Natural Science Foundation of China (Grant No. 82572327), Science and Technology Plan Joint Program of Liaoning Province of China (Natural Science Foundation - General Program, 2024-MSLH-596) 、 Science and Technology Program of Liaoning Province of China (Provincial Doctoral Research Startup Fund Program, 2024- BS-258) and National Natural Science Foundation of China (Grant No. 62506243).

D. Center-Specific Adapter

In deep convolutional neural networks, shallow convolutional layers primarily extract low-level image features such as edges, textures, and local contrast. These low-level features are highly sensitive to the hardware parameters of medical imaging acquisition devices, scanning protocols, and image reconstruction algorithms. Due to variations in equipment models, imaging parameters, and reconstruction methods across different medical centers, medical images exhibit systematic shifts in their low-level statistical characteristics. This leads to significant inter-domain differences in the distribution of shallow-layer features within the model. Traditional normalization methods like Batch Normalization employ linear affine transformations for feature normalization, which struggle to adequately adapt to such complex cross-center distribution variations. To address this issue, we propose a Center-Specific Adapter for the shallow layers of the encoder. The CSA architecture is illustrated in Fig. 4. This module replaces the conventional BN layer with a Center-Specific Multi-Layer Perceptron (CenterSpecificMLP, CSMLP), designed to model and adapt to the systematic differences in low-level image statistics across centers. Specifically, for each medical center participating in training, we maintain an independent MLP. During forward propagation, input features are first standardized, followed by extraction of global channel feature vectors. Based on the center affiliation of each sample, the feature vector is routed to its corresponding MLP to generate center-specific feature adjustments. The core advantage of the center-specific MLP lies in its nonlinear modeling capability. Unlike standard BN, which adjusts feature distributions solely through linear scaling and bias, the MLP can learn complex nonlinear mapping functions. This enables simultaneous adaptation to statistical differences across multiple dimensions among different centers.

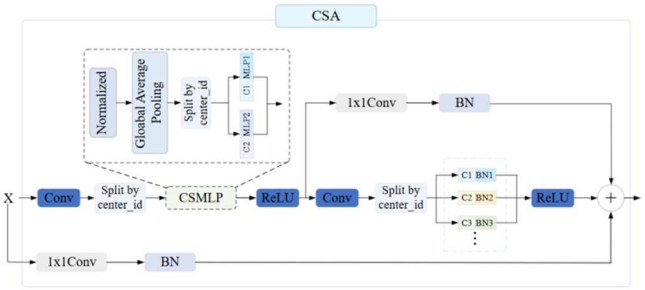

Fig. 4. Center-Specific Adapter

E. Dual-Path Feature Fusion Module

To achieve efficient multi-center data sharing and specificity modeling, we propose a dual-path feature fusion module implementing a dual-bottleneck layer architecture. The DPFFM structure is illustrated in Fig. 5. This module consists of two submodules: a shared bottleneck layer and a cross-center bottleneck layer, which are adaptively fused via learnable weights. The shared bottleneck layer adopts a classical convolutional block paradigm, comprising two 3×3 convolutional layers followed by a Batch Normalization layer. Although this module has a relatively large number of parameters, it is fully shared across all centers to learn a general anatomical structure representation across centers/datasets. This shared design significantly improves parameter efficiency, avoiding the storage and computational overhead of independent modeling for each center. The cross-center bottleneck layer focuses on modeling distribution variations across centers, centered around a low-rank cross-center transformation module. This module applies a low-rank transformation to the feature maps of each center, decomposing the transformation matrix into the product of two low-rank matrices, which significantly reduces the number of center-specific parameters. The low-rank transformation is followed by two lightweight 1×1 convolutions. The dual bottleneck layer is applied at the encoder's end, preserving the shared representational power of deep features while enabling flexible adaptation to multi-center distribution variations with minimal parameters. The outputs from both submodules are combined through a learnable fusion weight:

$$W = \alpha \cdot W_1 + (1-\alpha) \cdot W_2 \tag{7}$$

Where α is constrained to the interval [0,1] via the sigmoid function. W_1 denotes the output of the shared bottleneck layer, while W_2 represents the output of the cross-center bottleneck layer. This design enables the model to dynamically adjust the relative importance of general information versus center-specific information according to task requirements.

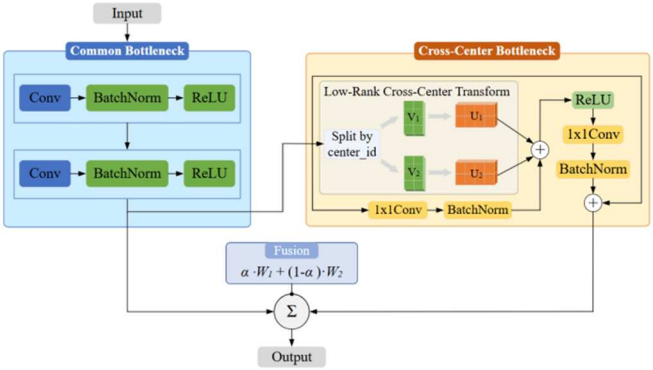

Fig. 5. Dual-Path Feature Fusion Module

F. Loss Function

This study employs a hybrid loss function combining Dice loss with a weighted binary cross-entropy loss (BCE).

1) Total loss function. The total loss function is defined as follows:

$$L_{\text{total}} = \lambda \cdot L_{\text{dice}} + (1-\lambda) \cdot L_{\text{bce}} \tag{8}$$

Where L_{dice} represents the Dice similarity coefficient loss, L_{bce} denotes the binary cross-entropy loss weighted by positive samples, and $\lambda = 0.7$ serves as the trade-off parameter.

2) Dice loss function. The Dice loss function is defined as:

$$L_{\text{dice}} = 1 - \frac{2\sum_i p_i g_i + \epsilon}{\sum_i p_i + \sum_i g_i + \epsilon} \quad (9)$$

Where p_i denotes the probability that the i-th pixel is predicted as tumor, g_i represents the corresponding ground truth label, and ϵ is the smoothing factor, which prevents the denominator from becoming zero and ensures numerical stability.

3) Weighted Binary Cross-Entropy Loss Function. The weighted binary cross-entropy loss is defined as:

$$L_{\text{bce}} = -\frac{1}{N}\sum_i \left[w \cdot g_i \cdot \log(\sigma(x_i)) + (1-g_i) \cdot \log(1-\sigma(x_i)) \right] \quad (10)$$

Where x_i denotes the raw output from the model's final layer, $\sigma(\cdot)$ represents the sigmoid activation function, and $w = 2.0$ is the positive sample weighting coefficient used to mitigate the class imbalance issue where tumor voxels (positive samples) are far fewer than background voxels (negative samples).

G. Parameter Complexity Analysis

To systematically evaluate the parameter efficiency of the proposed model in multi-center scenarios, we theoretically analyze the relationship between the model parameter scale and the number of centers C.

1) Overall Parameter Scale. The total number of parameters in MCA-UNet can be expressed as:

$$P_t = P_{\text{shared}} + P_{\text{center}} \cdot C \quad (11)$$

Where P_{shared} denotes the shared backbone parameters across all centers, including the encoder, decoder, and the common bottleneck; P_{center} represents the additional parameters introduced per center; and C is the number of centers. This formulation implies that the overall model complexity grows linearly with the number of centers.

2) Composition of Center-Specific Parameters. The additional parameters associated with each center are primarily derived from the following three components:

a) Center-Specific Batch Normalization Layer. Assume that there are l such normalization layers in the network, and the number of channels in the i-th layer is d_i. Then, the number of parameters for each center across all BN layers is:

$$P_B = \sum_{i=1}^{l} 2\, d_i \quad (12)$$

b) CenterSpecificMLP. In the CSA module, each center maintains an independent MLP. Let the channel dimension be d, and the hidden dimension be $(r \cdot d)$, where r is the reduction ratio. Then, the number of parameters of the MLP for each center is:

$$P_M = d\times(r\cdot d) + (r\cdot d)\times d = 2rd^2 \quad (13)$$

c) Cross-Center Bottleneck Layer. A low-rank cross-center transformation module is introduced in this bottleneck layer. Let the channel dimension be d and the decomposition rank be k. Then, the number of parameters for each center is:

$$P_L = 2\times d\times k + 2\times d = 2d(k+1) \quad (14)$$

which includes the low-rank decomposition matrices as well as the center-specific scaling and bias parameters.

By aggregating the above three components, the total number of additional parameters for each center is:

$$P_{\text{center}} = P_B + P_M + P_L \quad (15)$$

IV. Experiment

A. Experimental Setup

1) Dataset. This study evaluates model performance using the publicly available dataset from the HECKTOR 2021 Challenge [24]. This dataset comprises 3D PET/CT scan images of head and neck tumors. PET and CT images provide complementary information for tumor lesion segmentation at the metabolic activity and anatomical structure levels, respectively. The complete dataset comprises 224 cases from five distinct medical institutions (CHGJ, CHMR, CHUM, CHUS, and CHUP). Each case includes CT images, PET images, GTVt (Gross Tumor Volume) contours, corresponding bounding box coordinates, and patient-related metadata. For simplicity, this study selected data from two centers (CHGJ and CHMR) and employed a five-fold cross-validation strategy to divide the data into training and validation sets. Data preprocessing primarily involved the following steps: First, initial registration of CT and PET images was performed using 3D Slicer software; Subsequently, all images were resampled to a uniform size (128×128×64). For CT values, we truncated them to the range [-200, 200]. For PET images, we converted pixel values to SUV. Finally, both CT and PET images underwent Z-score normalization and were concatenated as input to the neural network.

2) Implementation Details. All models were trained for 100 epochs on an NVIDIA RTX 4090 GPU, with a batch size of 2 for the training set and 1 for the test set. The network employed AdamW [25] as the optimizer. The learning rate scheduling strategy combines CosineAnnealingLR [26] and ReduceLROnPlateau [27], with the minimum learning rate and decay factor set to 1e-5 and 0.5, respectively.

3) Evaluation metrics. In this study, we employed the Dice Similarity Coefficient (DSC) and 95% Hausdorff distance (HD95) to assess segmentation performance.

B. Comparative Experiment

1) Quantitative Comparison: In this paper, we compared the proposed MCA-UNet model with four deep learning models, including U-Net [4], AttUNet [28], UNet++ [18], and Swin-UNETR [29]. We modified these methods to accommodate dual-channel PET/CT inputs and trained and evaluated them under identical experimental settings. We computed and compared the DSC and HD95 metrics, listing the results in TABLE I. As shown in TABLE I, the proposed MCA-UNet method demonstrates superior performance, achieving the highest average DSC (81.50) and HD95 (5.48 mm) across five-

fold cross-validation. Furthermore, our method outperformed all competitors in DSC across all folds and in HD95 for folds 1, 3, and 5. TABLE II shows the parameter counts for the comparison methods and MCA-UNet, revealing that our approach achieves performance gains while increasing parameters by only 2.66%. More importantly, adding one additional center increases parameters by merely 0.05%, representing a negligible increase.

TABLE I. Quantitative comparison of tumor segmentation performance on the HECKTOR 2021 Dataset. The best result is indicated in bold. P-values were obtained from paired Student's t-tests comparing each baseline method with MCA-UNet. *p<0.05, **p<0.01

Model	DSC(%)						HD95(mm)						p-value
	Fold1	Fold2	Fold3	Fold4	Fold5	Mean	Fold1	Fold2	Fold3	Fold4	Fold5	Mean	
U-Net[4]	80.71	78.92	80.34	81.81	77.16	79.79	5.45	5.84	9.04	**7.63**	8.30	7.25	0.021*
AttUNet[28]	78.33	78.09	81.38	80.76	73.82	78.48	4.82	**4.98**	4.57	8.26	8.40	6.21	0.020*
UNet++[18]	78.11	77.23	81.01	81.09	71.21	77.73	4.90	6.14	8.95	8.94	5.32	6.85	0.037*
SwinUNETR[29]	72.19	75.76	78.73	78.08	75.70	76.09	14.09	6.51	6.69	18.63	5.65	10.31	0.008**
MCA-UNet	**81.94**	**79.34**	**83.32**	**83.19**	**79.70**	**81.50**	**4.69**	5.58	**4.13**	8.46	**4.54**	**5.48**	-

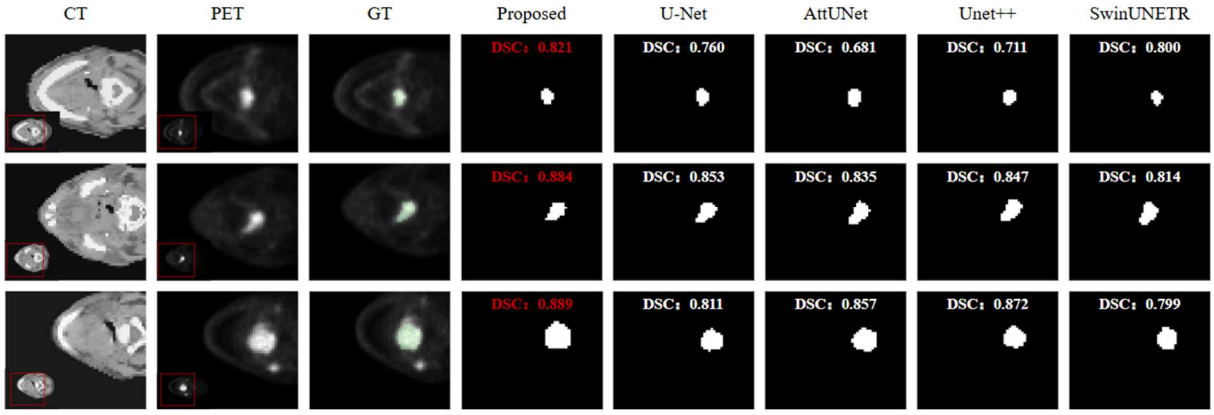

Fig. 6. Qualitative comparison of tumor segmentation results on the HECKTOR 2021 dataset. Green shading represents the ground truth. The Dice similarity coefficient is used to evaluate and compare the segmentation performance of different methods on the samples.

2) Qualitative Comparison: Fig. 6 presents CT images, PET images, ground truth, and segmentation results from MCA-UNet, U-Net, AttUNet, UNet++, and SwinUNETR for three representative samples. Visual comparisons reveal that our proposed MCA-UNet consistently achieves higher DSC scores than all other models, demonstrating optimal and most stable segmentation accuracy.

TABLE II. Model parameters. Where "+ΔParams (%)" denotes the percentage increase in parameters relative to the baseline U-Net model.

Model	Center Number	Params/M	+ΔParams (%)
U-Net	2	40.16	0.00%
MCA-UNet	2	41.23	2.66%
MCA-UNet	3	41.25	2.71%

C. Ablation Study

To validate the effectiveness of our proposed modules, we conducted ablation studies using U-Net as a baseline. The experimental results are shown in TABLE III. Among them, U-Net-DAB denotes replacing the standard convolutions in both the encoder and decoder with DAB modules on the U-Net architecture; U-Net-MBCCB denotes replacing the encoder DAB modules with MBCCB modules (one MBCCB module replaces two DAB modules) on the U-Net-DAB architecture; U-Net-MBCCB-CSA denotes the model where the first MBCCB layer in the encoder of U-Net-MBCCB is replaced with a CSA module; MCA-UNet is the final proposal of this paper. As shown in TABLE III, the baseline standard U-Net exhibits the lowest performance. Directly introducing the DAB module into U-Net (U-Net-DAB) improves DSC from 79.79% to 80.15%. This improvement demonstrates that the center specificity provided by DAB lays the foundation for subsequent multi-center adaptive refinement. The middle three rows of Table III show that introducing both the MBCCB and CSA modules yields performance gains, validating the effectiveness of MBCCB and CSA respectively. Finally, we evaluate the impact of the complete multi-center adaptive framework. As shown in the fifth row of TABLE III, the complete MCA-UNet model integrating all components (MBCCB, CSA, and DPFFM) achieved optimal performance. This result strongly confirms that our proposed multi-center collaborative learning framework is effective, with each component collectively forming a synergistically enhanced system. In summary, ablation

experiments systematically validated the effectiveness of the proposed modules and the overall framework.

D. Analysis of Fusion Weights

To explore the model's learning preference between shared features and center-specific features, we further analyzed the fusion weight α in the DPFFM module. As shown in Fig. 7, α consistently converged to approximately 0.62 in all five folds, indicating that shared anatomical features play a dominant role in tumor segmentation, while center-specific adaptation serves a supplementary and refined function.

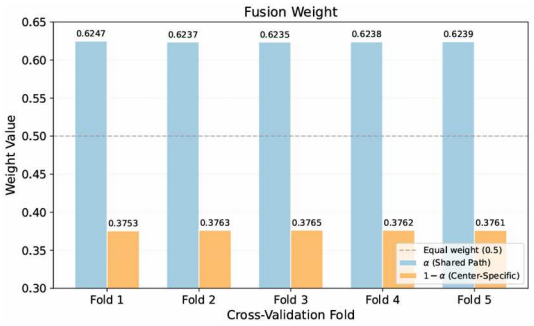

Fig. 7. Final distribution of fusion weight α in the DPFFM module.

TABLE III. Ablation study. Evaluate the contributions of the MBCCB, CSA, and DPFFM modules.

Model	DSC(%)
U-Net	79.79
U-Net-DAB	80.15
U-Net-DAB-MBCCB	80.31
U-Net-DAB-MBCCB-CSA	80.70
MCA-UNet	**81.50**

V. Discussion

A. Validity Analysis

In this study, we propose a multi-center adaptive U-Net architecture—MCA-UNet—for head and neck tumor segmentation in multi-center PET/CT images. By incorporating CSA, MBCCB, DAB, and DPFFM, MCA-UNet effectively mitigates the impact of inconsistent multi-center data distributions on model performance while maintaining a modest increase in parameters. Experimental results demonstrate that the proposed MCA-UNet achieves superior segmentation performance compared to other mainstream models on the HECKTOR2021 dataset. We attribute this improvement primarily to the following three factors:

1) MCA-UNet incorporates CSA into shallow encoder layers to adapt to low-level image variations across centers (e.g., noise, contrast, artifacts); deeper layers employ MBCCB to address semantic-level feature shifts. This hierarchical center adaptation mechanism aligns with the principle that features at different levels within deep networks exhibit varying sensitivities to distribution shifts.

2) The dual residual design in MBCCB preserves cross-center universal features (such as anatomical structures) while allowing center-specific features to be adapted through center-specific BN. This design enables the model to share knowledge across different centers while maintaining adaptability to specific distributions.

3) The DPFFM module achieves efficient fusion of generic features and center-specific features through the processing of a shared feature path and a low-rank cross-center path. The learning of the fusion weight α enables the model to dynamically balance the importance of these two components, adapting to the requirements of different centers or different cases.

B. Limitations

Although MCA-UNet performs well in multi-center segmentation tasks, it still has the following limitations:

1) Scalability across centers: This study only validated the model using data from two centers. While the module design supports more centers (as shown by limited parameter growth in TABLE II), its performance in scenarios with more centers and greater distribution variability requires further validation.

2) Modality combination limitations: Current experiments are designed for PET/CT dual-modality scenarios. Adaptability to other modality combinations (e.g., MRI/PET) or single-modality scenarios remains untested.

VI. Conclusion

This paper addresses the issue of data heterogeneity in multi-center medical image segmentation by proposing a Multi-center Adaptive UNet (MCA-UNet). The performance gains of the proposed method are mainly attributed to the following innovations: a hierarchical center adaptation mechanism that introduces center-specific adapters in the shallow encoder to capture low-level imaging variations; a multi-branch cross-center block in the deeper layers to enable both generic feature reuse and center-specific adaptation; and a dual-path feature fusion module in the bottleneck layer, which leverages the collaboration between a shared path and a low-rank cross-center path to achieve a parameter-efficient balance between generic representation learning and center-specific adaptation. Experiments on the public HECKTOR2021 head and neck tumor segmentation dataset demonstrate that MCA-UNet achieves significant performance improvements over existing methods while introducing only approximately 1.07M additional parameters. This study provides a solution for cross-center medical image segmentation that balances both performance and parameter efficiency.

References

[1] H. Xue, Y. Yao, and Y. Teng, "M-AECA Net: A Mamba-Based Auxiliary Encoder with Cross-Attention Fusion Network for PET/CT Tumor Segmentation," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–11, 2025..

[2] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of positron emission tomography images," *Computers in Biology and Medicine*, vol. 50, pp. 76–96, 2014.

[3] I. Shiri, M. Amini, F. Yousefirizi, P. Geramifar, M. R. Ay et al., "Information fusion for fully automated segmentation of head and neck tumors from PET and CT images," *Medical Physics*, vol. 51, no. 1, pp. 319–333, 2024.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, Cham, Switzerland: Springer International Publishing, 2015, pp. 234–241.

[5] M. Wang, H. Jiang, T. Shi, and Y. -d. Yao, "HD-RDS-UNet: Leveraging Spatial-Temporal Correlation Between the Decoder Feature Maps for Lymphoma Segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 1116–1127, 2022.

[6] J. Mei, C. Lin, Y. Qiu, Y. Wang, H. Zhang, et al., "Cross-Modal Interactive Perception Network with Mamba for Lung Tumor Segmentation in PET-CT Images," in 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2025, pp. 15653-15662.

[7] W. Zhao, Z. Huang, S. Tang, W. Li, Y. Gao, Y. Hu, et al., "MMCA-NET: A Multimodal Cross Attention Transformer Network for Nasopharyngeal Carcinoma Tumor Segmentation Based on a Total-Body PET/CT System," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 9, pp. 5447-5458, Sept. 2024

[8] M. J. Willemink, W. A. Koszek, C. Hardell, J. Wu, D. Fleischmann et al., "Preparing medical imaging data for machine learning," *Radiology*, vol. 295, no. 1, pp. 4–15, 2020.

[9] A. Kujawa, R. Dorent, S. Connor, A. Ourselin, and T. Vercauteren, "Deep learning for automatic segmentation of vestibular schwannoma: A retrospective study from multi-center routine MRI," *Frontiers in Computational Neuroscience*, vol. 18, p. 1365727, 2024.

[10] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati et al., "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," *Scientific Reports*, vol. 10, p. 12598, 2020.

[11] L. Wang, D. Li, H. Liu, S. Zhu, and Y. Wang, "Cross-dataset collaborative learning for semantic segmentation in autonomous driving," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 3, pp. 2487–2494, 2022.

[12] J. D. L. Araujo, L. B. da Cruz, J. O. B. Diniz, A. C. Silva, A. C. de Paiva et al., "Liver segmentation from computed tomography images using cascade deep learning," *Computers in Biology and Medicine*, vol. 140, p. 105095, 2022.

[13] X. Bai, Y. Hu, G. Gong, Y. Hao, and F. Wang, "A deep learning approach to segmentation of nasopharyngeal carcinoma using computed tomography," *Biomedical Signal Processing and Control*, vol. 64, p. 102246, 2021.

[14] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[15] J. Ni, J. Shi, Q. Zhan, Z. Zhang, and Y. Gu, "An improved U-Net model with multiscale fusion for retinal vessel segmentation," *Intelligent Robotics*, vol. 5, no. 3, pp. 679–694, 2025.

[16] N. Li, M. He, A. Mei, Z. Wu, T. Zhang et al., "Ranking-Aware Predictor-Assisted Automatic Architecture Design for Biomedical Image Segmentation,"in 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Wuhan, China, 2025, pp. 1-8.

[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431-3440.

[18] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in International Workshop on Deep Learning in Medical Image Analysis (DLMIA), Granada, Spain, Sep. 2018, pp. 3–11.

[19] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, Cham, Switzerland: Springer, Oct. 2016, pp. 424–432.

[20] Q. Hou, P. Cao, J. Wang, J. Cheng, Y. Yang et al., "Self-supervised domain adaptation for breaking the limits of low-quality fundus image quality enhancement," 2023, arXiv:2301.06943. [Online]. Available: https://arxiv.org/abs/2301.06943

[21] V. Cheplygina, M. de Bruijne, and J. P. W. Pluim, "Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Medical Image Analysis*, vol. 54, pp. 280–296, 2019.

[22] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon et al., "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2531–2540, 2020.

[23] S. Wang, L. Yu, K. Li, X. Yang, C.-W. Fu et al., "Dofe: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4237–4248, 2020.

[24] V. Andrearczyk, V. Oreiller, S. Boughdad, M. H. J. A. Eekers, L. F. S. P. Freitas et al., "Overview of the HECKTOR challenge at MICCAI 2021: automatic head and neck tumor segmentation and outcome prediction in PET/CT images," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*, Cham, Switzerland: Springer International Publishing, Sep. 2021, pp. 1–37.

[25] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, arXiv:1711.05101. [Online]. Available: https://arxiv.org/abs/1711.05101

[26] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," 2016, arXiv:1608.03983. [Online]. Available: https://arxiv.org/abs/1608.03983

[27] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[28] O. Oktay, J. Schlemper, L. L. Folgoc, M. J. Lee, M. P. Heinrich et al., “Attention u-net: learning where to look for the pancreas,” 2018, arXiv:1804.03999. [Online]. Available: https://arxiv.org/abs/1804.03999

[29] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth et al., "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in International MICCAI Brainlesion Workshop, Cham, Switzerland: Springer International Publishing, Sep. 2021, pp. 272–284.