Global-Local nnUNet with Anatomical-Metabolic Statistical Priors for Improved PET/CT based Whole Body Lymphoma Segmentation

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

2nd Yarong Feng
School of Electronic and Information Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
18093490172@163.com

3rd Yongwei Tang
School of Electronic and Information Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
2458672708@qq.com

4th Man Li
School of Electronic and Information Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
3375988726@qq.com

5th Yuxin Liang
School of information Science and Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
2833391387@qq.com

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

7th Tianyu Shi*
School of information Science and Engineering
Shenyang Ligong University
Shenyang city, 110159, P.R.China
ty.shi@sylu.edu.cn

***Abstract*—Automatic segmentation methods for lymphoma based on whole-body positron emission tomography (PET) and computed tomography (CT) still face significant challenges due to metabolic overlap between physiologically highly metabolizing and excreting organs (sFEPUs) and lesions. To address this, this paper proposes a whole-body lymphoma segmentation method that integrates anatomical and metabolic statistical priors. First, the CT-based whole-body organ segmentation method TotalSegmentator is employed to automatically segment 107 tissue/organ types across 300 PET/CT datasets. The metabolic intensity distribution of each tissue/organ is statistically analyzed to construct a population-level human metabolic statistical model. Building upon this foundation, an organ-specific metabolic suppression strategy is introduced for physiologically hypermetabolic regions, thereby reducing the complexity of the whole-body lymphoma segmentation task. Furthermore, a two-stage global-to-local segmentation framework is established: nnUNet performs coarse segmentation across the entire body, followed by local PET/CT-based fine segmentation using nnUNet to refine segmentation based on the probability map generated in the first stage, which locates potential lesion regions. Experimental results on the AutoPET dataset demonstrate that this method effectively reduces false positives caused by physiologically hypermetabolic organs and outperforms state-of-the-art medical image segmentation methods in overall segmentation performance.**

Keywords—Whole-body lymphoma segmentation, PET/CT, Anatomical-metabolic statistic prior, Global-Local Network, nnUNet

I. INTRODUCTION

Whole-body positron emission tomography/computed tomography (PET/CT) is a critical imaging modality for lymphoma assessment, staging, and treatment response evaluation, and has been widely incorporated into clinical practice guidelines [1-3]. PET reflects tissue metabolic activity, while CT provides precise anatomical information, enabling whole-body PET/CT to deliver both functional and structural data in clinical settings. In PET imaging, the level of metabolic activity is typically semi-quantitatively assessed using the standardized uptake value (SUV), which normalizes radioactive tracer uptake to facilitate comparisons across different tissues, individuals, and scanning conditions. Lymphoma lesions usually manifest as areas of increased 18F-FDG uptake; however, certain normal organs such as the brain, heart, liver, kidneys, and bladder also exhibit consistently high metabolic characteristics. These physiological high-metabolism and high-excretion regions (sFEPU) [4] overlap significantly with lesions in terms of metabolic intensity, complicating the automated segmentation of whole-body lymphoma and leading to a high propensity for false positives in models. Furthermore, lymphoma lesions are widely distributed throughout the body, exhibiting significant heterogeneity in size, morphology, and metabolic activity, which demands that models simultaneously possess the ability to comprehend global context and capture local details.

Currently, deep learning-based automated segmentation methods, particularly general segmentation frameworks like nnU-Net [5], have achieved significant progress in medical image analysis. However, when directly applied to whole-body PET/CT lymphoma segmentation tasks, their performance remains markedly constrained [6-8]. Existing approaches predominantly rely on feature fusion or contextual modeling capabilities within deep neural networks, with limited explicit incorporation of prior constraints for sFEPUs. This forces models to learn lesion features against highly complex backgrounds with significant metabolic noise, compromising segmentation accuracy and generalization capabilities [9].

* Corresponding author

Furthermore, traditional strategies based on fixed SUV thresholds can preliminarily label high-metabolic regions but fail to account for metabolic variations across organs and individuals. This often leads to misclassifying sFEPUs as lesions or missing lesions in their vicinity [10-12]. Bi et al.[4] employed patient-specific liver SUV reference values for whole-body SUV thresholding and utilized superpixel aggregation of sFEPUs to identify their organ affiliation. However, extracted sFEPUs typically exhibit fragmented distributions, making it challenging to form stable and discriminative representations compared to end-to-end learning directly from raw images. Consequently, relying solely on existing deep learning models or SUV thresholding methods remains insufficient for achieving accurate and robust whole-body lymphoma segmentation [13-15].

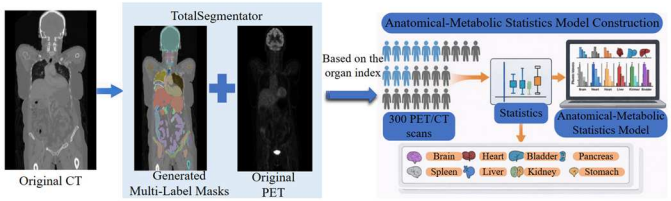

Fig.1. Workflow for constructing the anatomical-metabolic statistical model based on TotalSegmentator organ segmentation

To address these challenges, this study proposes an automated segmentation method for whole-body lymphoma that integrates anatomical-metabolic statistical priors with a global-local two-stage learning strategy. First, we employed TotalSegmentator to segment 300 PET/CT cases from the AutoPET dataset, comprising 200 healthy individuals and 100 lymphoma patients. This dataset exhibits no overlap with subsequent AutoPET datasets used in our experiments. We segmented 107 tissue/organ structures from these data, including multiple physiologically highly metabolized organs (sFEPUs) such as the brain, heart, liver, both kidneys, and bladder [16]. Subsequently, we performed statistical analysis of metabolic intensity (SUV) across anatomical structures, calculating metrics like mean and standard deviation for each organ to construct a population-level human metabolic statistical model. This model provides a basis for quantifying normal metabolic variation. In this paper, we focus solely on statistical analysis of sFEPUs; subsequent studies will further explore metabolic characteristics of other anatomical regions. Second, based on this model, we propose an organ-specific sFEPU suppression strategy. This approach adaptively determines metabolic filtering thresholds for the brain, heart, liver, kidneys, and bladder while detecting abnormal metabolic regions, thereby reducing interference from normal high-metabolism areas in training data. Subsequently, we designed a global-local two-stage nnU-Net framework: Stage 1 trains a global nnU-Net for coarse segmentation on sFEPU-suppressed whole-body images; Stage 2 first locates contiguous tumor regions using Stage 1's probability maps, then constructs corresponding local 3D image-label pairs by integrating CT, PET, and annotated data, finally performing fine segmentation with a local nnU-Net. This design enables the model to synergistically leverage global context and local multimodal information for precise optimization. Key contributions of this study include:

• Establishing an anthropometric model based on anatomical metabolic priors, providing quantitative justification for physiologically high-metabolism organs (sFEPUs).

• Proposed an organ-specific sFEPU suppression strategy that adaptively sets metabolic thresholds for different organs based on statistical priors, effectively reducing false positives and simplifying segmentation tasks.

• Designed a global-local two-stage nnU-Net segmentation framework that combines whole-body coarse segmentation with local fine-grained modeling to achieve high-precision, robust segmentation of systemic lymphoma.

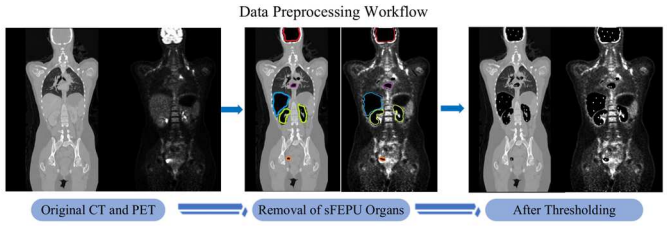

Fig. 2. Schematic diagram of sFEPU suppression

II. Related Works

A. PET/CT Lymphoma Automatic Segmentation

Automatic segmentation of lymphoma lesions in whole-body PET/CT scans is a critical task in clinical image analysis. Early methods primarily relied on SUV thresholding strategies or clustering approaches [10-11,17] for segmenting FDG-PET images, but struggled to account for inter-organ metabolic variations. Hatt et al. [18] evaluated the limitations of various thresholding methods in lymphoma segmentation, while Weisman et al. [19] further compared 11 automated segmentation approaches, confirming traditional methods' difficulty in reconciling lesions with complex anatomical backgrounds. With deep learning advancements, 3D U-Net [20] and its variants became mainstream. Chen et al. [21] developed a deep learning segmentation framework for whole-body lymphoma. To fully leverage PET/CT multimodal information, Song et al. [22] proposed a dual-path fusion network architecture processing distinct modal features, while Guo et al. [23] introduced cross-modal attention mechanisms to dynamically adjust feature weights. Wang et al. [24] proposed an anatomy-metabolism consistency generative adversarial network (AMC-GAN) approach to enhance the capture of metabolic abnormality appearance (MAA) through unsupervised learning. However, most of the above methods rely on network architecture improvements or single threshold rules, failing to systematically address the interference of normal hypermetabolic regions on model learning from both anatomical and metabolic perspectives.

B. Integration of Prior Knowledge in Medical Image Segmentation

In medical image segmentation tasks, incorporating prior knowledge—such as anatomical spatial information, shape constraints, and topological structures—into deep learning models has been demonstrated to enhance segmentation performance and robustness, particularly when lesions are

poorly defined or structures are complex. Dalca et al. [25] employed generative probabilistic models combined with anatomical priors for unsupervised segmentation, while Clough et al. [26] utilized topological priors to ensure shape consistency in predicted segments. Additional approaches embed statistical atlases or shape models into loss functions to assist networks in learning structural features [27-28]. In specific PET imaging scenarios, studies have also combined anatomical spatial information with CNN models to reduce interference from normal tissues with high metabolism (e.g., the bladder) on tumor segmentation [29]. Although these approaches attempt to integrate prior information at different levels—shape, topology, spatial distribution—most rely on generic anatomical or structural attributes. They have not yet established organ-specific prior constraints tailored to metabolic behavior, such as physiologically hypermetabolic regions in PET.

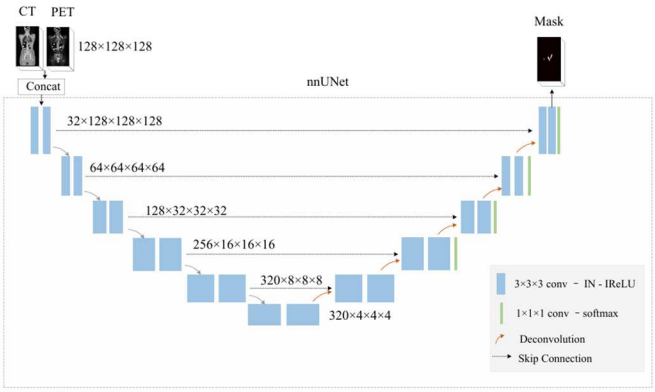

Fig. 3. Coarse segmentation network architecture for whole-body lymphoma

C. Multi-Stage Global-Local Segmentation Strategy

To address the challenge of balancing global context and local details in medical image segmentation, multi-stage or cascaded segmentation strategies have emerged as a key research direction. Such methods typically perform coarse localization or feature extraction at the global level, followed by local fine segmentation on candidate regions. This approach enhances the ability to identify small-scale targets, edge details, and complex backgrounds. Roth et al. [30] proposed a coarse-to-fine cascading strategy on 3D FCNs, significantly improving segmentation accuracy for small organs in CT organ segmentation. PL Net [31] and RefineSeg [32] further enhanced robustness for weakly supervised or complex structure segmentation through staged feature refinement or coarse-to-fine joint training. Recent studies have also employed hybrid CNN-Transformer architectures to achieve parallel modeling of global and local features, improving recognition capabilities for complex tumor or organ structures [33-34]. However, existing methods primarily rely on fixed geometric constraints or heuristic rules for region selection. They cannot dynamically adjust the focus area based on the specific anatomical and metabolic information of the current sample, leading to wasted computational resources and bottlenecks in segmentation accuracy for challenging regions.

III. Methodology

A. Human Anatomy – Construction of Metabolic Statistics Prior Models

To systematically quantify the anatomical-metabolic characteristics of whole-body PET/CT and provide robust prior information for subsequent lymphoma segmentation, we constructed a population-level metabolic statistical model covering the entire body's anatomical structures. This model employs multi-level statistical analysis from individual subjects to the population level, characterizing the metabolic distribution patterns of various organs and tissues in healthy individuals. This lays the foundation for distinguishing physiologically hypermetabolic regions from pathological lesions. The specific workflow is illustrated in Fig. 1.

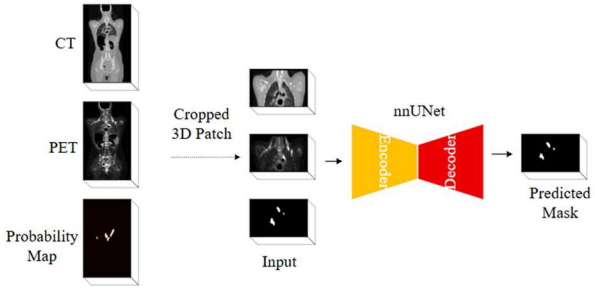

Fig. 4. Global context-guided local refinement segmentation network workflow, where the nnUNet model architecture is identical to that in Fig 3.

1) Data Preparation and Anatomical Segmentation: To construct a stable and representative human metabolic statistical model, we employed an independent cohort of 300 whole-body PET/CT scans from the AutoPET dataset, comprising 200 healthy individuals and 100 lymphoma patients. The demographic characteristics of this cohort cover a wide age range (22–75 years) and diverse sex distribution (165 males, 135 females), which well reflects typical clinical population distributions and thereby supports the generalization capability of the proposed population-level hierarchical prior. This cohort is completely non-overlapping with the AutoPET dataset used in subsequent experiments, avoiding any bias in model evaluation due to prior information leakage.

All PET and CT data undergo rigorous spatial registration. Specifically, we utilize the ANTs (Advanced Normalization Tools) registration framework within the 3D Slicer platform to align PET images to their corresponding CT space, achieving spatial consistency between the two modalities at the voxel level. The aligned PET images are represented in CT reference space as body-weight–normalized standardized uptake value (SUVbw). This approach ensures consistency in both spatial and scale representation between metabolic information and anatomical structures, thereby meeting the requirements for cross-patient statistical analysis.

For anatomical structure acquisition, we applied the TotalSegmentator[35] automatic segmentation algorithm to all CT images, yielding voxel-level segmentation results for the entire body's anatomical structures. This algorithm reliably covers major organs and tissue structures while demonstrating robust generalization performance across different scanning

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) and Science and Technology Program of Liaoning Province of China (Provincial Doctoral Research Startup Fund Program, 2024- BS-258) and the National Natural Science Foundation of China (Grant No. 62506243).

protocols and imaging conditions. Subsequently, the anatomical segmentation results were directly mapped onto the registered PET space. This approach enabled the extraction of metabolic information within corresponding regions under well-defined anatomical prior knowledge, laying the foundation for subsequent organ-level statistical analysis.

2) Single-Subject Anatomical Structure SUV Statistics： For any anatomical structure k in a subject's PET image, we define its PET voxel set as S_k . Based on S_k , we calculate the mean SUV (1) and standard deviation (2) for structure k:

$$\mu_k=\frac{1}{|S_k|}\sum_{v\in S_k} SUV(v) \tag{1}$$

Here, v denotes a voxel within structure k in the PET image, and $SUV_{(v)}$ represents the standardized uptake value of voxel v.

$$\sigma_k=\sqrt{\frac{1}{|S_k|}\sum_{v\in S_k}\left(SUV(v)-\mu_k\right)^2} \tag{2}$$

3) Construction of Population-Level Metabolic Statistics Profiles: After obtaining anatomical structure statistics for each subject, we aggregate them to form population-level metabolic statistics profiles. For structure k, its population mean μ_k^g(3) and population standard deviation σ_k^g (4) are defined as:

$$\mu_k^g=\mathrm{E}_i\left[\mu_k(i)\right] \tag{3}$$

$$\sigma_k^g=\mathrm{E}_i[\sigma_k(i)] \tag{4}$$

Here, i denotes different subjects, while $\mu_k(i)$ and $\sigma_k(i)$ represent the mean and standard deviation of subject i in structure k respectively. E_i denotes the mean across all subjects.

B. Organ-Specific sFEPU Threshold Based on Statistical Priors

Based on the aforementioned anatomy-metabolism-based human statistical model, we further propose an organ-specific sFEPU suppression strategy. This method explicitly suppresses physiological uptake regions in six categories of normally highly metabolizing and excreting organs—the brain, heart, liver, both kidneys, and bladder—during the data preprocessing stage. This reduces background complexity in whole-body PET/CT images, facilitating more discriminative representation for subsequent lymphoma segmentation network learning.

1) sFEPU Organ Localization and Anatomical Constraints: To achieve precise suppression of physiologically highly metabolized organs (sFEPU), reliable localization of relevant organs within the target dataset (AutoPET) is first required. We similarly perform PET-CT registration and represent the registered PET images in SUVbw format. Building upon this foundation, we applied the TotalSegmentator algorithm to segment each CT scan. Based on the official anatomical structure indexing system defined by TotalSegmentator, we extracted six key sFEPU organs from the segmentation results: the brain, heart, liver, left kidney, right kidney, and bladder. These serve as anatomical spatial constraints for subsequent metabolic normalization and suppression processing.

To further enhance the stability and usability of organ masks, we performed simple morphological post-processing on the automated segmentation results, including small connected component removal and hole filling. This approach eliminates potential segmentation noise and ensures spatial continuity of anatomical structures. The resulting sFEPU organ masks provide reliable anatomical constraints for subsequent organ-specific metabolic normalization and suppression based on statistical priors.

2) sFEPU Organ Metabolic Normalization: Following sFEPU organ localization, we utilize the previously constructed anatomical-metabolic human statistical model to normalize PET voxels within each organ. This process eliminates differences in absolute SUV values across organs, yielding a comparable metric. This approach was primarily inspired by the work of Christian et al. [36]. In that study, the authors proposed utilizing Z-scores to statistically summarize whole-body metabolic information. Specifically, for any voxel v within organ k , its normalized standard score (Z-score) is defined as (5):

$$Z_k(v)=\frac{SUV(v)-\mu_k^g}{\sigma_k^g} \tag{5}$$

Through normalization, voxel metabolic intensity can be compared with the metabolic levels of healthy individuals in the same organ, providing a foundation for subsequent organ-specific threshold determination.

After completing Z-score normalization for organ-specific metabolites, we established adaptive absolute value thresholds for each organ k to effectively distinguish normal physiological uptake from abnormally high metabolic activity. This approach draws inspiration from [37-38], which employed fixed Z-score thresholds as demarcation points. Consequently, we defined the organ-specific threshold as (6):

$$T_k=\mu_k^g+Z_k(\mathrm{v})\cdot\sigma_k^g \tag{6}$$

We selected Z=3 as the threshold for detecting regions with significant SUV abnormalities. Furthermore, statistical analysis of SUV distributions from 300 PET/CT scans yielded thresholds reflecting normal metabolic information for different organs, with specific values shown in Table I.

When $SUV_v\leq T_k$, voxel v is considered a normal physiological metabolic voxel. When $SUV_v>T_k$, it indicates a metabolic intensity significantly higher than the normal range for organs in the general population, suggesting a potential pathological lesion.

C. Global-to-Local Two-Stage nnUNet Segmentation：

To simultaneously capture both global contextual information across the entire body and local tumor details, we designed a global-local two-stage segmentation strategy based on nnUNet, aiming to achieve high-precision segmentation of systemic lymphoma：

1) sFEPU Suppression Module: To explicitly mitigate interference from physiologically hypermetabolic organs (sFEPUs) during model training and construct a network suitable for whole-body lymphoma segmentation. This study designed a preprocessing workflow for CT and PET images and their corresponding gold standards in the AutoPET dataset, based on organ priors and metabolic statistical characteristics. The specific workflow is shown in Fig. 2. This workflow performs targeted suppression of physiologically

hypermetabolic backgrounds at the data level, ultimately generating multimodal inputs for the global segmentation network (7):

$$X_{\text{global}} = \left[C_{\text{remove}}, \tilde{P}\right], Y_{\text{clean}} \tag{7}$$

C_{remove} denotes CT images after sFEPU suppression, while $\tilde{P}$ represents PET images with simplified metabolic background. Y_{clean} indicates processed labels. Specifically, based on the sFEPU organ binary mask M_k(which $k \in \{$ brain,heart,liver, kidney$_{\text{left}}$,kidney$_{\text{right}}$,bladder$\}$) obtained in Section III.B.1, an adaptive threshold T_k is used to filter voxels within each organ region in the PET image space. voxels with $SUV_v \leq T_k$ are classified as background to be suppressed, forming the background mask M_k^b . Given that the PET and CT images have been strictly spatially aligned during acquisition, the background mask M_k^b defined in the PET space is directly mapped to the aligned CT image space. Ultimately, for voxels in the PET image that fall within any background mask M_k^b, their SUV values are set to 0. For voxels in the CT image that fall within any background mask M_k^b, their CT values are uniformly set to the standard CT value for air (-1024 HU), while the CT values in other regions remain unchanged.

$$C_{\text{remove}}(v) = \begin{cases} -1024 & \exists k \text{ s.t. } v \in M_k^b \\ C(v) & \text{otherwise.} \end{cases} \tag{8}$$

$$\tilde{P}(v) = \begin{cases} 0 & \exists k \text{ s.t. } v \in M_k^b \\ P(v) & \text{otherwise} \end{cases} \tag{9}$$

This operation significantly reduced the interference of these high-uptake organs on model training. Similarly, we applied corresponding consistency processing to the original tumor mask.

$$Y_{\text{clean}}(v) = \begin{cases} 0 & \exists k \text{ s.t. } v \in M_k^b \\ Y(v) & \text{otherwise.} \end{cases} \tag{10}$$

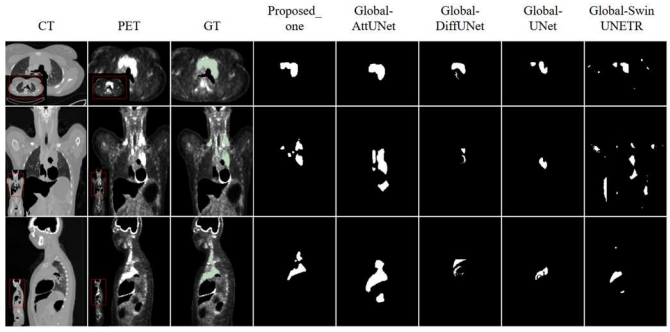

Fig. 5. Qualitative comparison of different methods in the global segmentation stage for PET/CT lymphoma segmentation task

2) Anatomy-Metabolism-Guided Coarse Segmentation Network for Whole-Body Lymphoma: In the first stage, we constructed a coarse segmentation network covering the entire body to generate preliminary lymphoma probability maps. These maps provide spatial prior information and confidence guidance for subsequent local refinement, as shown in Figure 3. This stage utilizes preprocessed X_{global} as input. To address the challenges of extremely low voxel coverage and highly imbalanced category distribution for lymphoma lesions in whole-body PET/CT images, our study adopted the default foreground oversampling strategy designed for sparse segmentation tasks within the nnUNet framework. During each training batch construction, image regions containing at least one foreground (lymphoma) voxel are selected for training with a 33% probability. This strategy significantly reduces the dominance of pure background samples in training, compelling the network to focus on learning tumor feature representations. Consequently, it effectively mitigates model bias caused by class imbalance while enhancing the network's sensitivity to sparsely distributed small lesions and segmentation stability.

3) Globally Context-Guided Local Refinement Segmentation Network: Since the whole-body segmentation network in the first stage achieves a relatively reliable overall segmentation of tumor regions, the high-confidence regions in the resulting global probability map exhibit a high true positive rate, thereby effectively mitigating the impact of false negatives on the subsequent local refinement stage. The local refinement module in the second stage utilizes the probability map generated by the whole-body segmentation network in the first stage to perform fine-grained segmentation on high-confidence tumor regions. To accurately locate the tumor regions requiring refinement within the global probability map, this study proposes an adaptive region of interest (ROI) extraction method based on axial continuity. Specifically, the global probability map undergoes maximum intensity projection along the Z-axis to obtain the maximum tumor confidence sequence for each slice. An empirical threshold τ=0.3 is applied to binarize this sequence, identifying candidate tumor slices. Three-dimensional connected component analysis based on axial continuity principles then delineates spatially contiguous tumor regions. To fully capture transitional information at tumor boundaries and surrounding context, each identified contiguous region undergoes boundary expansion along the Z-axis. The number of slices expanded upward and downward is set to 3, forming the final local regions of interest (ROIs).

For each local ROI, we crop the corresponding 3D region from the preprocessed PET/CT images, global probability maps, and labels, as shown in Figure 4. Note that to ensure training data validity and model stability, manually inspect the cropped 3D blocks and exclude those containing no tumor voxels to prevent invalid samples from interfering with network training.

IV. Experiments

A. AutoPET Dataset

The AutoPET[39] dataset serves as a comprehensive PET/CT tumor imaging benchmark, comprising 1,611 paired 3D CT and PET scans from multiple centers. This includes 1,014 FDG cases (from Tübingen University Hospital, Siemens Biograph mCT) and 597 PSMA tracer cases (from Munich University, Siemens Biograph 64 4R, Siemens Biograph mCT Flow 20, GE Discovery 690). Each paired CT/PET image is annotated with tumor lesions, covering multiple tumor types including malignant melanoma, lymphoma, and lung cancer, and providing heterogeneity across tracers, centers, and scanner vendors. CT volumes were resampled to match PET resolution and dimensions, and PET values were converted to

standardized uptake values (SUV) for improved interpretability. This dataset provides a challenging benchmark for developing models that handle multi-tracer and multi-center volumetric tumor segmentation. In our experiments, focusing on lymphoma segmentation tasks, we utilized a lymphoma subset comprising 136 PET/CT cases.

TABLE II. Dice coefficient of different methods in five-fold cross-validation. P-values obtained from paired Student's t-tests comparing each baseline method with the Proposed method were all below 0.05.

Model \ Fold	1	2	3	4	5	mean	p-value
Global-UNet	0.384	0.343	0.524	0.517	0.629	0.479	0.0175
Global-AttUNet	0.363	0.283	0.440	0.475	0.372	0.387	0.0008
Global-Swin UNETR	0.346	0.284	0.482	0.458	0.438	0.402	0.0011
Global-DiffUNet	0.419	0.457	0.492	0.403	0.402	0.435	0.0001
Local-UNet	0.472	0.621	0.499	0.532	0.468	0.518	0.0033
Local-AttUNet	0.330	0.532	0.470	0.384	0.319	0.407	0.0023
Local-Swin UNETR	0.453	0.346	0.348	0.407	0.301	0.371	0.0001
Local-DiffUNet	0.380	0.302	0.397	0.331	0.321	0.346	0.0001
Proposed_one	0.676	0.639	**0.707**	0.649	**0.671**	0.668	-
Proposed	**0.713**	**0.695**	0.681	**0.696**	0.657	**0.689**	-

B. Experimental Settings

For consistency and fair comparison, all experiments employ use 5-fold cross-validation to evaluate model performance and are conducted on NVIDIA RTX 4090 GPUs. Our proposed method is implemented based on the 3d_fullres[5] configuration of the nnUNet v2 framework. It leverages its automated pipeline to optimize the network topology for the dataset, including core hyperparameters such as encoder-decoder depth, convolution kernel size, and activation functions, thereby avoiding the subjective influence of manual parameter tuning. The experiments are divided into two stages: In the first stage, the global segmentation network is trained using 136 preprocessed PET/CT images. In the second stage, the local refinement network is trained based on 360 local 3D blocks extracted from the predictions of the first stage. All training employs an 8:2 split between training and validation sets, with Dice coefficient serving as the primary performance metric.

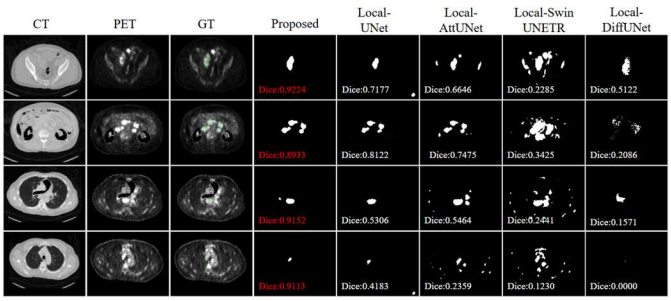

Fig. 6. Qualitative comparison of different methods during the local segmentation phase for PET/CT lymphoma segmentation tasks. Using the axial view as an example, from top to bottom: AutoPET_20_block00, AutoPET_20_block01, AutoPET_20_block02, AutoPET_20_block03.

C. Comparative Experiments

To validate the effectiveness and superiority of the proposed method in the task of whole-body PET/CT lymphoma segmentation, we systematically compared it with several representative segmentation methods, including the traditional convolutional neural network method U-Net [40], Attention U-Net [41] which introduces an attention mechanism, Swin UNETR [42], and DiffUNet [43].

1) Quantitative Results Analysis: Table II summarizes the Dice coefficient results of each method under five-fold cross-validation. For fairness, all local models were trained using the same three-dimensional image blocks processed with anatomical-metabolic priors as in the second stage of our method. Experimental results demonstrate that the proposed complete two-stage method achieves the optimal average Dice coefficient (0.689), significantly outperforming all baseline models. Even under identical data conditions, its performance substantially exceeds that of the best-performing local U-Net baseline (0.518), validating the effectiveness of the introduced anatomical-metabolic prior and the global-local two-stage strategy.

TABLE I. sFEPU threshold

Organ	μ_k^g	σ_k^g	T_k
Brain	7.98	2.08	14.22
Heart	2.54	1.16	6.02
Liver	2.08	0.42	3.34
Left Kidney	2.24	0.61	4.07
Right Kidney	2.36	0.60	4.16
Urinary Bladder	12.97	11.10	46.27

2) *Qualitative Results Analysis*: As shown in Fig. 5 and 6, we performed a comparative visualization and analysis of the segmentation results obtained by different methods at both the global and local fine-grained segmentation levels. It can be observed that conventional methods commonly suffer from issues such as under-detection, over-segmentation, or fragmented predictions. In contrast, the proposed method can more stably focus on the true lesion regions, and its predictions are highly consistent with the ground truth (GT) in terms of spatial location, morphological continuity, and boundary coherence. These observations indicate that incorporating anatomical-metabolic priors combined with a global-to-local two-stage strategy helps to improve the robustness and clinical applicability of the model in complex PET/CT segmentation tasks.

D. Ablation Study

To evaluate the effectiveness of each key module in the proposed method, we conducted ablation experiments with a stepwise introduction strategy. The results are shown in Table III. The baseline model, trained directly on raw PET/CT data using nnUNet without incorporating any anatomical-metabolic

priors or multi-stage designs, achieved an average Dice coefficient of 0.617. This indicates that single-stage models are susceptible to interference from high metabolic backgrounds in the presence of widespread physiological high-metabolism organs (sFEPU). Subsequently, the sFEPU suppression strategy based on anatomical-metabolic statistical priors was introduced, along with global coarse segmentation on preprocessed whole-body images (Table III： Proposed_one), resulting in an improvement of the average Dice to 0.668. This demonstrates that organ-specific metabolic suppression effectively reduces false positives and simplifies the segmentation task. Further integration of the global-local two-stage segmentation framework (Table III： Proposed), which utilizes global probability maps to guide fine segmentation of local 3D patches, continued to enhance model performance, achieving the best result with a Dice coefficient of 0.689. These experiments demonstrate that even without modifying the network architecture, the progressive introduction of anatomical-metabolic priors at both the data level and task modeling level, along with staged strategies, can significantly improve the accuracy and robustness of whole-body PET/CT lymphoma segmentation.

TABLE III. Model ablation results under different stage strategies

Model \ Fold	1	2	3	4	5	mean
nnUNet	0.702	0.633	0.582	0.563	0.605	0.617
Proposed_one	0.676	0.639	**0.707**	0.649	**0.671**	0.668
Proposed	**0.713**	**0.695**	0.681	**0.696**	0.657	**0.689**

V. Discussion

In this study, we introduced an anatomical-metabolic statistical prior and combined it with a global-local two-stage nnU-Net framework, effectively enhancing the accuracy and stability of whole-body PET/CT lymphoma segmentation. Through an organ-specific sFEPU suppression strategy, physiological high-metabolic interference during training and inference was significantly reduced, leading to superior performance in spatial localization, lesion connectivity, and boundary consistency. Qualitative visualization results further validated this strategy's ability to reduce false-positive predictions in complex anatomical regions. Although overall performance has markedly improved, limitations persist in the bladder region, primarily due to the mixed modeling of urine and bladder wall metabolic features. Future work will explore more refined structural segmentation for complex organs like the bladder and introduce organ-specific hierarchical metabolic priors to further enhance robustness and clinical applicability in challenging lesion segmentation areas.

VI. Conclusion

This paper proposes a global-local two-stage nnU-Net framework guided by anatomical-metabolic statistical priors for automatic segmentation of lymphoma in whole-body PET/CT. By constructing population-level anatomical-metabolic statistical models and introducing an organ-specific suppression strategy for physiologically hypermetabolic regions, it effectively reduces false positive interference from normally hypermetabolic organs and simplifies the segmentation task. Combined with a two-stage segmentation framework, it excels in both global localization and local detail refinement. Experimental results demonstrate that this method outperforms multiple comparative approaches in segmentation accuracy and false positive suppression, providing an effective solution for precise automated segmentation of systemic lymphoma.

References

[1] S. F. Barrington, N. G. Mikhaeel, L. Kostakoglu, M. Meignan, M. Hutchings, S. P. Müeller, L. H. Schwartz, E. Zucca, R. I. Fisher, J. Trotman *et al.*, "Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group," *Journal of Clinical Oncology*, vol. 32, no. 27, pp. 3048–3058, 2014.

[2] S. F. Barrington and R. Kluge, "FDG PET for therapy monitoring in Hodgkin and non-Hodgkin lymphomas," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. Suppl 1, pp. 97–110, 2017.

[3] R. Boellaard, R. Delgado-Bolton, W. J. Oyen, F. Giammarile, K. Tatsch, W. Eschner, F. J. Verzijlbergen, S. F. Barrington, L. C. Pike, W. A. Weber et al., "FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 42, no. 2, pp. 328–354, 2015.

[4] L. Bi, J. Kim, A. Kumar, L. Wen, D. Feng, and M. Fulham, "Automatic detection and classification of regions of FDG uptake in whole-body PET-CT lymphoma studies," *Computerized Medical Imaging and Graphics*, vol. 60, pp. 3–10, 2017.

[5] F. Isensee, P. F. Jaeger, S. A. A. Kohl *et al.*, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, pp. 203–211, 2021.

[6] C. S. Constantino, S. Leocádio, F. P. M. Oliveira, M. Silva, C. Oliveira, J. C. Castanheira, Â. Silva, S. Vaz, R. Teixeira, M. Neves *et al.*, "Evaluation of Semiautomatic and Deep Learning-Based Fully Automatic Segmentation Methods on [18F]FDG PET/CT Images from Patients with Lymphoma: Influence on Tumor Characterization," *Journal of Digital Imaging*, vol. 36, no. 4, pp. 1864–1876, 2023.

[7] M. E. Rayed, S. M. S. Islam, S. I. Niha, J. R. Jim, M. M. Kabir, and M. F. Mridha, "Deep learning for medical image segmentation: State-of-the-art advancements and challenges," *Informatics in Medicine Unlocked*, vol. 47, p. 101504, 2024.

[8] C. S. Constantino *et al.*, "A robust deep-learning model for fully automatic segmentation of lymphoma lesions on whole-body [18F]FDG PET/CT images," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, Athens, Greece, 2024, pp. 1–4.

[9] Y. Peng, J. Kim, D. Feng, and L. Bi, "Automatic Tumor Segmentation vi a False Positive Reduction Network for Whole-Body Multi-Modal PET/ CT Images," 2022, arXiv:2209.07705. [Online]. Available: https://arxiv.org/abs/2209.07705

[10] N. Mansour, J. Joram, F. Grawe, A. Hinterberger, J. Rübenthaler, K. Klambauer, W. G. Kunz, M. Winkelmann, C. C. Cyran, J. Ricke, O. Öcal, M. Unterrainer, K. Jurkschat, C. Wängler, B. Wängler, R. Schirrmacher, A. Nitschmann, T. Greve, G. Sheikh, A. Holzgreve, and M. P. Fabritius, "Validation of SUV thresholds in [^{18}F]SiTATE PET/CT for accurate meningioma segmentation," *Eur. J. Nucl. Med. Mol. Imaging*, 2025. doi: 10.1007/s00259-025-07476-9.

[11] D. M. L. Lilburn, H. S. Garthwaite, B. Ganeshan, T. Win, N. J. Screaton, L. R. Hoy, D. Walls, R. Endozo, R. I. Shortman, F. Fraioli, A. U. Wells, C. P. Denton, A. M. Groves, and J. C. Porter, "[18F]FDG PET/CT predicts patient survival in patients with systemic sclerosis-associated interstitial lung disease," *J. Nucl. Med.*, vol. 66, no. 7, pp. 1135–1141, Jul. 2025. doi: 10.2967/jnumed.125.269497.

[12] A. Samim, D. P. D. Suurd, R. van Rooij, M. M. van Noesel, M. G. E. H. Lam, A. J. A. T. Braat, N. Tolboom, L. Borgwardt, G. A. M. Tytgat, and B. de Keizer, "SUV normalisation and reference tissue selection for [^{18}F]mFBG PET-CT in paediatric and adult patients," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 52, no. 10, pp. 3773–3780, Oct. 2025. doi: 10.1007/s00259-025-07242-x.

[13] H. Yu, C. Caldwell, K. Mah, and D. Mozeg, "Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation tre

atment planning," *IEEE Transactions on Medical Imaging*, vol. 28, no. 3 , pp. 374–383, 2009.

[14] S. Vauclin, K. Doyeux, S. Hapdey, A. Edet-Sanson, P. Vera, and I. Gard in, "Development of a generic thresholding algorithm for the delineation of 18FDG-PET-positive tissue: application to the comparison of three thr esholding models," *Physics in Medicine and Biology*, vol. 54, no. 21, pp. 6901–6916, 2009.

[15] S. Gu, Z. Dai, L. Wu, W. Liu, and L. Chen, "Dempster-Shafer Theory E mpowered Deep Learning for Enhanced 3D PET-CT Medical Image Seg mentation: A first test on Automated Lesion Segmentation in Whole-bod y Dataset," in *2024 4th International Conference on Neural Networks, In formation and Communication Engineering (NNICE)*, 2024, pp. 350–35 4.

[16] J. Wasserthal, H. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. Boll, J. Cyriac, S. Yang *et al.*, "TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images," *Radiology: Artificial Intelligence*, 2023.

[17] C. M. Bregenzer, L. M. Knappe, A. Weissensee, N. Gözlügöl, A. Afshar-Oromieh, C. Mingels, C. Gräni, A. Rominger, and F. Caobelli, "New semiquantitative parameters in digital [18F]FDG PET/CT improve diagnostic accuracy in suspected infective endocarditis," *EJNMMI Rep.*, vol. 9, no. 1, p. 21, 2025. doi: 10.1186/s41824-025-00256-6.

[18] M. Hatt, F. Tixier, L. Pierce, P. E. Kinahan, C. C. Le Rest, and D. Visvikis, "Characterization of PET/CT images using texture analysis: the past, the present… any future?," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. 1, pp. 151–165, 2017.

[19] A. J. Weisman, M. W. Kieler, S. Perlman, M. Hutchings, R. Jeraj, L. Kostakoglu, and T. J. Bradshaw, "Comparison of 11 automated PET segmentation methods in lymphoma," *Physics in Medicine and Biology*, vol. 65, no. 23, p. 235019, 2020.

[20] O. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, LNCS, vol. 9901, Oct. 2016, pp. 424–432.

[21] X. Chen, S. Sun, N. Bai, K. Han, Q. Liu, S. Yao, H. Tang, C. Zhang, Z. Lu, Q. Huang *et al.*, "A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy," *Radiotherapy and Oncology*, vol. 160, pp. 175–184, 2021.

[22] J. Song, X. Yang, X. Liang, J. Huang, J. Ma, Y. Sun, W. Luo, S. Mok, Y. Wang, and T. Tan, "DuEU-Net: Dual Encoder UNet with Modality-Agnostic Training for PET-CT Multi-modal Organ and Lesion Segmentation," *Artificial Intelligence and Imaging for Diagnostic and Treatment Challenges in Breast Care.* Deep-Breath 2024, Springer, LNCS, vol. 15451, 2025.

[23] Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, "Deep learning-based image segmentation on multimodal medical imaging," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 162–169, 2019. doi: 10.1109/trpms.2018.2890359.

[24] 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.

[25] A. V. Dalca, J. Guttag, and M. R. Sabuncu, "Anatomical priors in convolutional networks for unsupervised biomedical segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 9290-9299. doi: 10.1109/CVPR.2018.00968.

[26] J. R. Clough, I. Oksuz, N. Byrne, J. A. Schnabel, and A. P. King, "Explicit topological priors for deep-learning based image segmentation using persistent homology," in *Inf. Process. Med. Imag. (IPMI)*, vol. 11492, A. Chung, J. Gee, P. Yushkevich, and S. Bao, Eds. Cham, Switzerland: Springer, 2019.doi: 10.1007/978-3-030-20351-1_2.

[27] H. Huang, H. Zheng, L. Lin, M. Cai, H. Hu, Q. Zhang, Q. Chen, Y. Iwamoto, X. Han, Y. W. Chen, and R. Tong, "Medical image segmentation with deep atlas prior," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3519–3530, 2021. doi: 10.1109/TMI.2021.3089661.

[28] P. Zhang, Y. Cheng, and S. Tamura, "Shape prior-constrained deep learning network for medical image segmentation," *Comput. Biol. Med.*, vol. 180, p. 108932, 2024. doi: 10.1016/j.compbiomed.2024.108932.

[29] L. Chen, C. Shen, Z. Zhou, G. Maquilan, K. Albuquerque, M. R. Folkert, and J. Wang, "Automatic PET cervical tumor segmentation by combining deep learning and anatomic prior," *Phys. Med. Biol.*, vol. 64, no. 8, p. 085019, 2019. doi: 10.1088/1361-6560/ab0b64.

[30] H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, and K. Mori, "An application of cascaded 3D fully convolutional networks for medical image segmentation," *Comput. Med. Imag. Graph.*, vol. 66, pp. 90–99, 2018. doi: 10.1016/j.compmedimag.2018.03.001.

[31] K. Mao, R. Li, J. Cheng, D. Huang, Z. Song, and Z. Liu, "PL-Net: progressive learning network for medical image segmentation," *Front. Bioeng. Biotechnol.*, vol. 12, p. 1414605, 2024. doi: 10.3389/fbioe.2024.1414605.

[32] A. Du, N. Aung, T. N. Arvanitis, S. K. Piechnik, J. A. C. Lima, S. E. Petersen, and L. Zhang, "RefineSeg: dual coarse-to-fine learning for medical image segmentation," 2025, arXiv:2508.02844. [Online]. Available: https://arxiv.org/abs/2508.02844

[33] J. Chen, Y. Lu, Q. Yu, and et al., "TransUNet: transformers make strong encoders for medical image segmentation," 2021, arXiv:2102.04306. [Online]. Available: https://arxiv.org/abs/2102.04306

[34] X. Liu, Y. Hu, and J. Chen, "Hybrid CNN-Transformer model for medical image segmentation with pyramid convolution and multi-layer perceptron," *Biomed. Signal Process. Control*, vol. 86, p. 105331, 2023. doi: 10.1016/j.bspc.2023.105331.

[35] J. Wasserthal, H. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, M. Bach, and M. Segeroth, "TotalSegmentator: robust segmentation of 104 anatomic structures in CT images," *Radiol. Artif. Intell.*, vol. 5, no. 5, p. e230024, 2023. doi: 10.1148/ryai.230024.

[36] K. Takeda, T. Hara, X. Zhou, T. Katafuchi, M. Kato, S. Ito, K. Ishihara, S. Kumita, and H. Fujita, "Normal model construction for statistical image analysis of torso FDG-PET images based on anatomical standardization by CT images from FDG-PET/CT devices," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 5, pp. 777–787, 2017. doi: 10.1007/s11548-017-1526-4.

[37] D. Hellwig, T. P. Graeter, D. Ukena, A. Groeschel, G. W. Sybrecht, H. J. Schaefers, and C. M. Kirsch, "18F-FDG PET for mediastinal staging of lung cancer: which SUV threshold makes sense?," *J. Nucl. Med.*, vol. 48, no. 11, pp. 1761–1766, 2007. doi: 10.2967/jnumed.107.044362.

[38] J. F. Vansteenkiste, S. G. Stroobants, P. R. De Leyn, P. J. Dupont, J. Bogaert, A. Maes, G. J. Deneffe, K. L. Nackaerts, J. A. Verschakelen, T. E. Lerut, L. A. Mortelmans, and M. G. Demedts, "Lymph node staging in non-small-cell lung cancer with FDG-PET scan: a prospective study on 690 lymph node stations from 68 patients," *J. Clin. Oncol.*, vol. 16, no. 6, pp. 2142–2149, 1998. doi: 10.1200/JCO.1998.16.6.2142.

[39] S. Gatidis, T. Hepp, M. Früh, C. La Fougère, K. Nikolaou, C. Pfannenberg, B. Schölkopf, T. Küstner, C. Cyran, and D. Rubin, "A whole-body FDG-PET/CT dataset with manually annotated tumor lesions," *Sci. Data*, vol. 9, no. 1, p. 601, 2022. doi: 10.1038/s41597-022-01718-3.

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[41] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, arXiv:1804.03999. [Online]. Available: https://arxiv.org/abs/1804.03999

[42] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. Roth, and D. Xu, "Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images," 2022, arXiv:2201.01266. [Online]. Available: https://arxiv.org/abs/2201.01266

[43] Z. Xing, L. Wan, H. Fu, G. Yang, Y. Yang, L. Yu, B. Lei, and L. Zhu, "Diff-UNet: A diffusion embedded network for robust 3D medical image segmentation," *Med. Image Anal.*, vol. 105, p. 103654, 2025. doi: 10.1016/j.media.2025.103654.