HWA-UNETR: Hierarchical Window Aggregate UNETR for 3D Multimodal Gastric Lesion Segmentation

Jiaming Liang[1] , Lihuan Dai[2] , Xiaoqi Sheng[1] , Xiangguang Chen[3], Chun Yao[3], Guihua Tao[4], Qibin Leng[5], Hongmin Cai[1] , and Xi Zhong[2] [(✉)]

[1] School of Computer Science and Engineering, School of Future Technology, South China University of Technology, Guangzhou, China
[2] Department of Medical Imaging, Guangzhou Institute of Cancer Research, the Affiliated Cancer Hospital, Guangzhou Medical University, Guangzhou, China
`zhongxi@gzhmu.edu.cn`
[3] Department of Radiology, Meizhou People's Hospital, Meizhou, China
[4] School of Computer Science and Technology, Hainan University, Hainan, China
[5] Department of Oncology Institute, Guangzhou Institute of Cancer Research, the Affiliated Cancer Hospital, Guangzhou Medical University, Guangzhou, China

Abstract. Multimodal medical image segmentation faces significant challenges in the context of gastric cancer lesion analysis. This clinical context is defined by the scarcity of independent multimodal datasets and the imperative to amalgamate inherently misaligned modalities. As a result, algorithms are constrained to train on approximate data and depend on application migration, leading to substantial resource expenditure and a potential decline in analysis accuracy. To address those challenges, we have made two major contributions: First, we publicly disseminate the GCM 2025 dataset, which serves as the first large-scale, open-source collection of gastric cancer multimodal MRI scans, featuring professionally annotated FS-T2W, CE-T1W, and ADC images from 500 patients. Second, we introduce HWA-UNETR, a novel 3D segmentation framework that employs an original HWA block with learnable window aggregation layers to establish dynamic feature correspondences between different modalities' anatomical structures, and leverages the innovative tri-orientated fusion mamba mechanism for context modeling and capturing long-range spatial dependencies. Extensive experiments on our GCM 2025 dataset and the publicly BraTS 2021 dataset validate the performance of our framework, demonstrating that the new approach surpasses existing methods by up to 1.68% in the Dice score while maintaining solid robustness. The dataset and code are public via this URL.

Keywords: Gastric Cancerous · GCM 2025 · Multimodal Segmentation.

1 Introduction

Gastric Cancer (GC) ranks as the fifth most common malignant neoplasm globally and the fourth leading cause of cancer-related mortality [23], whose detec-

tion of lesions is a critical step in clinical treatment. Through sustained technical refinements, Magnetic Resonance Imaging (MRI) has evolved into a superior modality for gastrointestinal malignancy evaluation [5], with its diagnostic ascendancy over Computed Tomography (CT) stemming from three cardinal virtues: multi-parametric soft tissue discrimination, absence of ionizing radiation, and synergistic incorporation of functional molecular imaging paradigms [22]. Recently, with the advancement of Deep Learning (DL) technologies, quantitative multimodal MRI scans have leveraged their inherent advantages to serve as an essential tool for detecting cancerous lesions [25,12,9]. More importantly, these technologies have significantly enhanced medical image segmentation methods [19,13], allowing for the systematic delineation of malignant tumor spatial distributions and precise quantification of lesion heterogeneity through the integration of multimodal data.

Existing DL methods are sensitive to imaging modalities and lesion characteristics, leading to segmentation performance that relies heavily on training data quality [18,4,1]. However, the scarcity of large-scale, high-quality multimodal MRI datasets for GC necessitate reliance on generic medical imaging datasets and transfer learning [16], both of which are resource-intensive and reduce specificity. Furthermore, gastric cancer lesion detection involves multiple spatially misaligned MRI modalities, which pose two critical challenges for DL methods: (1) **aligning inherently misaligned imaging modalities and integrating complementary features**, and (2) **fully leveraging the complementary characteristics of each modality for multi-scale context modeling**. Previous studies enhance segmentation via feature fusion [24,26], yet clinical applicability remains limited due to reliance on overly optimistic data assumptions. Additionally, improving model representational capacity has proven effective for segmentation tasks. CNN-based [11,20] and Transformer-based methods [9,13,19] improve contextual feature modeling, while Mamba-based methods [25,7] optimize the modeling process by incorporating state space models (SSMs). Nonetheless, these predominantly focus on pixel-level predictions within single modalities, neglecting multimodal misalignment and complementarity. Recently, H-DenseFormer [21] proposes a multi-path parallel embedding framework to enhance segmentation efficiency and accuracy in multimodal tasks. Meanwhile, MMEF [6] integrates Dempster-Shafer theory [6] with a DL network to achieve reliable feature extraction. Nevertheless, existing approaches still struggle to fully leverage the complementary features of gastric MRI modalities, limiting their effectiveness in achieving reliable lesion segmentation.

This study addresses challenges in multimodal MRI segmentation of gastric lesions with two principal contributions: (1) **GCM (Gastric Cancer MRI) 2025 dataset**, the first publicly available multimodal MRI resource for gastric cancer research, featuring fat-suppressed T2-weighted (FS-T2W), venous phase contrast-enhanced T1-weighted (CE-T1W), and apparent diffusion coefficient (ADC) images from 500 gastric cancer patients (as shown in Fig. 1), is released to provide essential baseline data for lesion segmentation; and (2) **HWA-UNETR** (**H**ierarchical **W**indow **A**ggregate **UNE**t **TR**ansformer), a 3D multimodal seg-
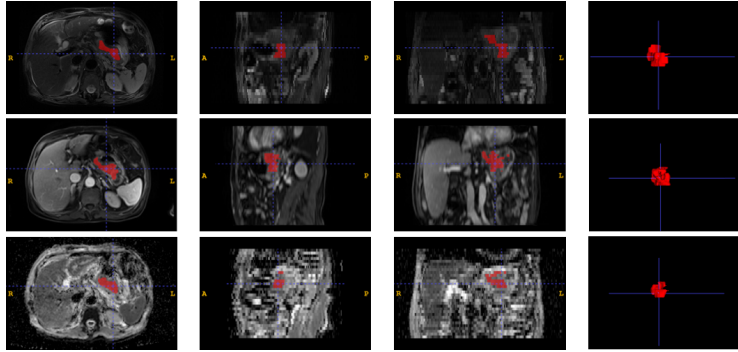
Fig. 1. The data visualization for GCM 2025 dataset. From top to bottom, the images demonstrate the FS-T2W, CE-T1W, and ADC modalities.

mentation framework is proposed, combining the novel HWA blocks for dynamic cross-modal feature correspondence and TFM blocks for global, multi-scale feature modeling through its innovative tri-orientated fusion mamba mechanism. Extensive experiments on multimodal benchmarks, including GCM 2025 and BraTS 2021 datasets, demonstrate the technical superiority of HWA-UNETR. The proposed framework achieves a Dice score improvement of up to 1.68% compared to State-Of-The-Art (SOTA) methods, while maintaining exceptional robustness across various organ structures and multimodal imaging scenarios.

2 GCM 2025: Gastric Cancer MRI Dataset

GCM (**G**astric **C**ancer **M**RI Dataset) 2025 is currently recognized as the first large-scale open-source dataset featuring professionally annotated gastric cancer MRI examinations, integrating multimodal MRI results from 500 gastric cancer patients. It establishes standardized baselines for multimodal DL segmentation techniques, thereby advancing epidemiological research on gastric cancer.

Dataset Modalities: The standardization protocol of GCM 2025 includes FS-T2W, CE-T1W, and ADC images, as illustrated in Fig. 1, where FS-T2W/CE-T1W demonstrate high soft-tissue contrast for lesion detection and tumor staging [3], while ADC-based diffusion imaging noninvasively quantifies water molecular mobility to characterize tumor pathophysiology [15].

Dataset Construction: MRI scans of GCM 2025 were acquired from November 2017 to December 2024. All examinations were performed using 1.5T or 3.0T scanners equipped with dedicated body coils. Sensitive patient information in this dataset has been removed. Each volume was annotated by a professional doctor and cross-checked by a second doctor. Detailed data acquisition information is provided at this public https URL.
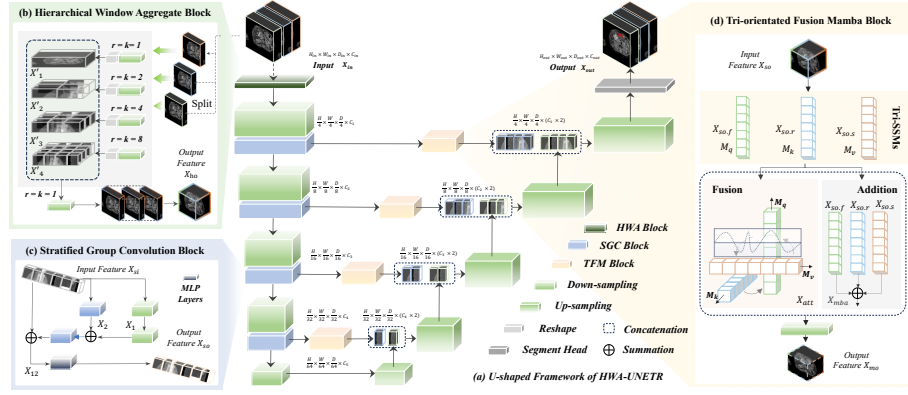
Fig. 2. Overview of HWA-UNETR.

3   Methodology

To address the challenges of multimodal gastric lesion segmentation, the HWA-UNETR framework is proposed, which achieves high-performance multimodal segmentation tasks through novel blocks and architecture design. The detailed implementation is illustrated in Fig. 2 and described in the following sections.

3.1   U-shaped Framework Overview

HWA-UNETR adopts a 4-stage U-shaped architecture comprising an encoder-decoder structure with skip connections, which is shown in Fig. 2 (a), aiming to achieve high-performance 3D multimodal image segmentation. The input volumetric data X_{in} with dimensions $H_{in} \times W_{in} \times D_{in} \times C_{in}$ (height, width, depth, modals) undergoes hierarchical feature extraction in the encoder. At each stage i, spatial resolution is progressively reduced to $\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times \frac{D}{2^{i+1}}$ via depthwise separable convolutional layers followed by instance normalization and ReLU activation. In the decoder, transposed convolutional layers upsample feature maps to recover the original spatial resolution. Simultaneously, skip connections integrate multi-scale encoder features to minimize information loss. Finally, a segmentation head equipped with a sigmoid activation layer produces pixel-wise probability maps, facilitating precise and scalable image segmentation.

To enhance its core functionality, HWA-UNETR integrates three key components within its framework: **H**ierarchical **W**indow **A**ggregate (HWA) block, **S**tratified **G**roup **C**onvolution (SGC) block, and **T**ri-orientated **F**usion **M**amba (TFM) block. These components are introduced in the following sections.

3.2   Hierarchical Window Aggregate Block

The HWA block, whose structure is shown in Fig. 2 (b), operates at the multimodal volumetric input X_{in} stage, dynamic window partitioning, and multi-kernel convolutions are employed to align cross-modal local semantics. This

structure allows for the extraction of features from multimodal images with cross-resolution characteristics, generating semantically consistent fused image features. Specifically, the HWA Block adaptively employs depthwise convolutional layers with varying strides r and kernel sizes k based on the input image shapes to extract features from multimodal images, as detailed in Eq. (1).

$$[X_1, X_2, ..., X_{C_{in}}] = Sp(X_{in});$$
$$X_i' = R(Dw(X_i)_1) \parallel R(Dw(X_i)_2) \parallel R(Dw(X_i)_4) \parallel R(Dw(X_i)_8); \quad (1)$$
$$X_{ho} = w_1 X_1' + w_2 X_2' + ... + w_{C_{in}} X_{C_{in}}';$$

where X_{ho} represents the output of the HWA Block; $+$ and $\parallel$ denote addition operation and concatenation operation performed along dimension C_{in}; w_i represents the learnable parameters; $Dw(.)_i$ represents the depthwise convolutional layer with $r = k = i$; R denotes the reshape operation that upsamples the convolution results back to their original size; and $Sp(.)$ represents a splitting operation that divides the input data into separate vectors via the C_{in} dimension.

3.3   Stratified Group Convolution Block

The SGC block, deployed post-downsampling in each encoder stage, enhances local feature retention and spatial information preservation for cross-level fusion [25]. As shown in Fig. 2 (c), the tri-branch architecture processes shared input X_{si} via: (1) dual depthwise convolutions (k=3,1) with instance normalization obtain X_1; (2) pointwise convolution refinement obtain X_2; and (3) residual-enhanced fusion with MLP transformation obtain X_{so} as result of SGC block.

3.4   Tri-orientated Fusion Mamba Block

The TFM Block is deployed in the skip connections stage, specifically designed for modeling global and multi-scale features, as shown in Fig. 2 (d).

In particular, to perform complementary feature modeling on cross-modal feature maps, the TFM block introduces a novel Tri-orientated Fusion Mamba computational mechanism, which integrates the long-sequence modeling capability of Tri-SSMs [25] with the cross-modal attention mechanisms inherent in the attention [9] mechanism. This mechanism, as mathematically formulated in Eq. (2), provides enriched feature representations specifically optimized for the decoder's upsampling operations.

$$M_q, M_k, M_v = Ma(X_{so.f}), Ma(X_{so.r}), Ma(X_{so.s});$$
$$X_{mba} = M_q + M_k + M_v;$$
$$X_{att} = S(\frac{M_q M_k^T}{\sqrt{d_{M_k}}})M_v; \quad (2)$$
$$X_{mo} = Dw(X_{mba} \parallel X_{att})_1;$$

where X_{mo} denotes the output of TFM Block; S denotes the softmax function [8], d_{M_k} denotes the dimensionality of M_k, Ma refers to the Mamba module that models global information within the sequence, $X_{so.f}$, $X_{so.r}$, and $X_{so.s}$ denote the forward, reverse, and inter-slice direction feature maps, respectively.

4 Experiments

4.1 Assistance Dataset

BraTS 2021 dataset [2] serves as an accessorial benchmark for evaluating HWA-UNETR's performance in multimodal MRI segmentation, particularly assessing its generalizability across tumor subregions. It contains 1,251 3D brain MRI cases, each comprising four imaging modalities (T1, T1Gd, T2, and T2-FLAIR) with annotations for three distinct tumor compartments: Whole Tumor (WT), Enhancing Tumor (ET), and Tumor Core (TC).

4.2 Implementation Details

Our model is implemented in PyTorch 2.4.1, CUDA 12.4, and MONAI 1.3.2, trained with random crops of 128×128×64 (GCM 2025) and 128×128×128 (BraTS 2021), batch size 2 per GPU. Both experiments utilized the AdamW optimizer for 300 epochs with an initial learning rate of 0.001, incorporating linear warm-up, cosine annealing, a weight decay of 0.4, and 1:1 positive-to-negative sample balancing, while employing a composite loss function that integrates soft dice loss [17] and focal loss [14]. All data underwent random axis flipping (50% prob.) and intensity scaling/shifting (20% prob.). Experiments run on a cloud platform with four NVIDIA A100 GPUs, randomly allocating 3D volumes into 70% training, 10% validation, and 20% testing.

4.3 Performance

We compare HWA-UNETR with five SOTA segmentation methods: CNN-based nnUNet [11], Transformer-based Swin UNETR [9], SSM-based SegMamba [25] / nnMamba [7], and MMEF [10], a method specifically designed for multimodal segmentation. All methods were retrained using their official implementations under consistent settings for fair comparison. Performance was quantitatively evaluated on GCM 2025 and BraTS 2021 datasets using Dice coefficient (Dice) and 95% Hausdorff Distance (HD95).

GCM 2025 :

Table 1 summarizes the multimodal segmentation results of HWA-UNETR and other SOTA methods on the newly released GCM 2025 dataset. The outstanding segmentation performance demonstrated in the results validates the effectiveness of our approach. Although HWA-UNETR ranks second only to the method MMEF [10] in terms of the optimal segmentation edge metric HD95 (5.03 mm), it achieves the highest average Dice score of 74.21%, surpassing

Table 1. Comparative Performance Analysis of Various Network Architectures for Lesion Segmentation on the GCM 2025 Dataset. The performance metrics of HWA-UNETR are highlighted in gray , whereas the best segmentation performance is marked in **red**.

Method	Dice Similarity Coefficient (%)				Average Hausdorff
	FS-T2W↑	CE-T1W↑	ADC↑	**Avg.**↑	Distance 95% (mm) ↓
nnUNet [11]	70.47	71.68	67.31	69.82	6.54
Swin-UNETR [9]	72.06	72.46	69.83	71.45	5.67
SegMamba [25]	71.6	73.83	71.05	72.16	5.37
nnMamba [7]	70.12	73.02	69.94	71.03	5.86
MMEF-nnUNet [10]	72.66	73.69	71.24	72.53	**5.03**
HWA-UNETR(ours)	**74.76**	**75.56**	**72.31**	**74.21**	5.35

Table 2. Comparative Performance Analysis of Various Network Architectures for Lesion Segmentation on the BraTS 2021 Dataset. The colors of the record in this table are labeled the same as in Table 1.

Method	Dice Similarity Coefficient (%)				Average Hausdorff
	WT↑	ET↑	TC↑	**Avg.**↑	Distance 95% (mm) ↓
nnUNet [11]	92.72	88.34	91.39	90.84	5.33
Swin-UNETR [9]	93.32	89.08	91.69	91.36	5.03
SegMamba [25]	93.13	88.21	93.82	91.72	4.73
nnMamba [7]	92.52	86.68	93.89	91.03	5.33
MMEF-nnUNet [10]	**93.58**	89.04	93.77	92.13	3.97
HWA-UNETR(ours)	93.41	**89.54**	**94.07**	**92.34**	**3.51**

other SOTA methods by 1.68%. This performance highlights HWA-UNETR's superior capability in handling multimodal segmentation tasks with misaligned information.

BraTS 2021 : The results on the BraTS 2021 dataset are listed in Table 2. The proposed HWA-UNETR achieves optimal performance compared to other methods, achieving an average Dice score of 92.34% along with an average HD95 of 3.51 mm. These compelling results further substantiate two key advantages of HWA-UNETR: (1) its advanced feature integration capabilities across diverse imaging modalities, and (2) its remarkable robustness in cross-organ segmentation tasks.

Visual Comparisons : Fig. 3 provides an intuitive comparison of the segmentation results obtained by different methods. On the GCM 2025 dataset, our HWA-UNETR accurately detects the boundaries of tumor regions in each modality image, demonstrating better consistency compared to other SOTA methods.

4.4   Ablation Study

Table 3 validates the effectiveness of core modules in the HWA-UNETR framework. On the GCM 2025 dataset: (1) The TFM Block establishes excellent performance (i.e, 71.84% Dice score) through global context modeling and hierar-
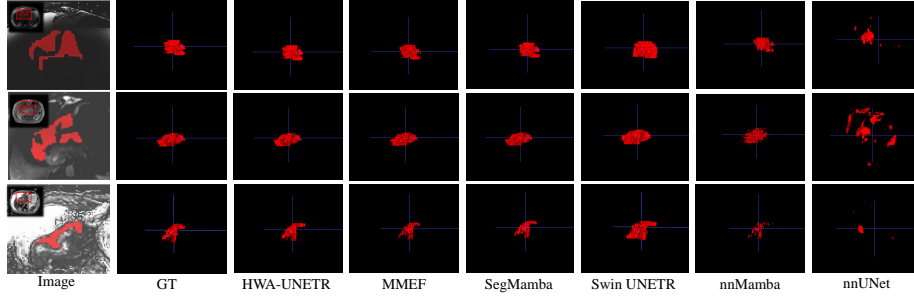
Fig. 3. Visual comparisons of proposed HWA-UNETR and other SOTA methods. The arrangement of the images in this figure corresponds to that in Fig. 1.

Table 3. Ablation study of various block settings of HWA-UNETR on the GCM 2025 benchmark. The colors of the record in this table are labeled same as in Table 1.

HWA	SGC	TFM	Avg. Dice↑	Avg. HD 95 ↓
×	×	×	67.56	10.57
×	×	✓	71.84	6.03
×	✓	✓	72.36	**5.31**
✓	✓	✓	**74.21**	5.35

chical feature integration; (2) the SGC Block enhances Dice score to 72.47% by capturing local information; and (3) the HWA Block achieves a 1.85% Dice score improvement (74.21%) via dynamic window partitioning strategy, demonstrating superior capability in handling misaligned multimodal data.

5   Conclusion

In this paper, HWA-UNETR is proposed as a novel multimodal 3D medical image segmentation framework specifically tailored for gastric cancer lesions. The framework designs a hierarchical window aggregation HWA Block with a learnable cross-modal alignment layer to address the structural differences between unaligned MRI modalities. In addition, a new tri-orientated fusion mamba is introduced to model global and multi-scale features for multimodal data specifically. To advance gastric cancer research, GCM 2025, the first large-scale open-source multimodal MRI dataset, is publicly released, comprising 500 expert-annotated, aligned multimodal MRI scans (FS-T2W/CE-T1W/ADC) for gastric cancer segmentation. Extensive experiments on both the newly released GCM 2025 and the public BraTS 2021 datasets demonstrate that HWA-UNETR achieves SOTA segmentation accuracy on gastric multimodal MRI scans while exhibiting exceptional generalizability in cross-organ segmentation tasks, outperforming conventional approaches by a significant margin.

Disclosure of Interests. This study and its authors have no competing interests.

References

1. Ali, M., Wu, T., Hu, H., Luo, Q., Xu, D., Zheng, W., Jin, N., Yang, C., Yao, J.: A review of the segment anything model (sam) for medical image analysis: Accomplishments and perspectives. Computerized Medical Imaging and Graphics p. 102473 (2024)
2. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021)
3. Bammer, R.: Basic principles of diffusion-weighted imaging. European Journal of Radiology **45**(3), 169–184 (2003)
4. Bibars, M., Salah, P.E., Eldeib, A., Elattar, M.A., Yassine, I.A.: Cross-modality deep transfer learning: Application to liver segmentation in ct and mri. In: Annual Conference on Medical Image Understanding and Analysis. pp. 96–110. Springer (2023)
5. Borggreve, A.S., Goense, L., Brenkman, H.J., Mook, S., Meijer, G.J., Wessels, F.J., Verheij, M., Jansen, E.P., Van Hillegersberg, R., Van Rossum, P.S., et al.: Imaging strategies in the management of gastric cancer: current role and future potential of mri. The British Journal of Radiology **92**(1097), 20181044 (2019)
6. Denœux, T., Dubois, D., Prade, H.: Representations of uncertainty in artificial intelligence: Probability and possibility. A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning pp. 69–117 (2020)
7. Gong, H., Kang, L., Wang, Y., Wan, X., Li, H.: nnmamba: 3d biomedical image segmentation, classification and landmark detection with state space model. arXiv preprint arXiv:2402.03526 (2024)
8. Han, D., Ye, T., Han, Y., Xia, Z., Pan, S., Wan, P., Song, S., Huang, G.: Agent attention: On the integration of softmax and linear attention. In: European Conference on Computer Vision. pp. 124–140. Springer (2024)
9. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI Brainlesion Workshop. pp. 272–284. Springer (2022)
10. Huang, L., Ruan, S., Decazes, P., Denœux, T.: Deep evidential fusion with uncertainty quantification and reliability learning for multimodal medical image segmentation. Information Fusion **113**, 102648 (2025)
11. Isensee, F., Jäger, P.F., Full, P.M., Vollmuth, P., Maier-Hein, K.H.: nnu-net for brain tumor segmentation. In: International MICCAI Brainlesion Workshop. pp. 118–132. Springer (2020)
12. Liang, J., Huang, T., Li, D., Ding, Z., Li, Y., Huang, L., Wang, Q., Zhang, X.: Agilenet: A rapid and efficient breast lesion segmentation method for medical image analysis. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). pp. 419–430. Springer (2023)

13. Liang, J., Zhang, M., Tan, C., Huang, T., Zhang, X., Zhang, Z., Gao, S., Sheng, Q., Pang, Y.: Comprehensive transformer integration network (ctin): Advancing endoscopic disease segmentation with hybrid transformer architecture. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). pp. 210–224. Springer (2024)
14. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2980–2988 (2017)
15. Liu, S., Guan, W., Wang, H., Pan, L., Zhou, Z., Yu, H., Liu, T., Yang, X., He, J., Zhou, Z.: Apparent diffusion coefficient value of gastric cancer by diffusion-weighted imaging: correlations with the histological differentiation and lauren classification. European Journal of Radiology **83**(12), 2122–2128 (2014)
16. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
17. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)
18. Pang, Y., Li, Y., Huang, T., Liang, J., Wang, Z., Dong, C., Kuang, D., Hu, Y., Chen, H., Lei, T., et al.: Online self-distillation and self-modeling for 3d brain tumor segmentation. IEEE Journal of Biomedical and Health Informatics (2025)
19. Pang, Y., Liang, J., Huang, T., Chen, H., Li, Y., Li, D., Huang, L., Wang, Q.: Slim unetr: Scale hybrid transformers to efficient 3d medical image segmentation under limited computational resources. IEEE Transactions on Medical Imaging (2023)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
21. Shi, J., Kan, H., Ruan, S., Zhu, Z., Zhao, M., Qiao, L., Wang, Z., An, H., Xue, X.: H-denseformer: An efficient hybrid densely connected transformer for multimodal tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 692–702. Springer (2023)
22. Singh, A., Adams-Tew, S., Johnson, S., Odeen, H., Shea, J., Johnson, A., Day, L., Pessin, A., Payne, A., Joshi, S.: Treatment efficacy prediction of focused ultrasound therapies using multi-parametric magnetic resonance imaging. In: MICCAI Workshop on Cancer Prevention through Early Detection. pp. 190–199. Springer (2024)
23. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians **71**(3), 209–249 (2021)
24. Wenxuan, W., Chen, C., Meng, D., Hong, Y., Sen, Z., Jiangyun, L.: Transbts: Multimodal brain tumor segmentation using transformer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 109–119 (2021)
25. Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 578–588. Springer (2024)
26. Zhou, T., Ruan, S., Canu, S.: A review: Deep learning for medical image segmentation using multi-modality fusion. Array **3**, 100004 (2019)