MBA-Net: SAM-driven Bidirectional Aggregation Network for Ovarian Tumor Segmentation

Yifan Gao[1,2], Wei Xia[2], Wenkui Wang[3], and Xin Gao[2]✉

[1] School of Biomedical Engineering (Suzhou), Division of Life Science and Medicine, University of Science and Technology of China, Hefei, China
[2] Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China
[3] State Key Laboratory of Ultra-precision Machining Technology, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, PR China
yifangao@mail.ustc.edu.cn
Wang.wenkui@polyu.edu.hk
xiaw@sibet.ac.cn, xingaosam@163.com

Abstract. Accurate segmentation of ovarian tumors from medical images is crucial for early diagnosis, treatment planning, and patient management. However, the diverse morphological characteristics and heterogeneous appearances of ovarian tumors pose significant challenges to automated segmentation methods. In this paper, we propose MBA-Net, a novel architecture that integrates the powerful segmentation capabilities of the Segment Anything Model (SAM) with domain-specific knowledge for accurate and robust ovarian tumor segmentation. MBA-Net employs a hybrid encoder architecture, where the encoder consists of a prior branch, which inherits the SAM encoder to capture robust segmentation priors, and a domain branch, specifically designed to extract domain-specific features. The bidirectional flow of information between the two branches is facilitated by the robust feature injection network (RFIN) and the domain knowledge integration network (DKIN), enabling MBA-Net to leverage the complementary strengths of both branches. We extensively evaluate MBA-Net on the public multi-modality ovarian tumor ultrasound dataset and the in-house multi-site ovarian tumor MRI dataset. Our proposed method consistently outperforms state-of-the-art segmentation approaches. Moreover, MBA-Net demonstrates superior generalization capability across different imaging modalities and clinical sites.

Keywords: Ovarian Tumor Segmentation · Medical Image Segmentation · Segment Anything Model · Bidirectional Aggregation Network.

1 Introduction

Ovarian cancer is one of the most lethal gynecological malignancies, accounting for over 200,000 deaths globally each year [1]. Early and accurate detection and

delineation of ovarian tumors is crucial for improving prognosis, treatment planning, and survival rates. However, the manual segmentation of ovarian tumors from medical images is a tedious, time-consuming, and subjective process, often leading to high inter- and intra-observer variability [2,3]. This has motivated the development of automated computational methods for ovarian tumor segmentation, which can provide objective and reproducible results, thereby assisting clinicians in making more informed decisions and ultimately improving patient care [4,5].

Traditional segmentation methods like graph cuts and level sets rely heavily on hand-crafted features and prior assumptions about tumor appearance, limiting their performance and generalization capability [6,7]. Recent years have seen a paradigm shift towards deep learning for medical image segmentation. Convolutional neural networks (CNN) and Transformers have achieved state-of-the-art performance by automatically learning hierarchical representations directly from data [8–12].

Nevertheless, automatic ovarian tumor segmentation has persisted as an open challenge, largely due to the significant diversity across ovarian cancer subtypes [13,14]. Over 30 histological classifications have been documented, with tumors exhibiting substantial variance in morphology, growth patterns, and imaging phenotypes [15]. Irregular tumor shape and ambiguous tumor boundaries further complicate segmentation tasks, especially in ultrasound [16], which is heavily utilized for ovarian cancer screening and diagnosis. Additionally, the lack of data on some relatively rare subtypes has constrained model performance and generalizability across imaging modalities and clinical centers [17]. These factors lead to suboptimal performance for complex subtypes like high-grade serous carcinomas and mucinous cysts, potentially impacting staging accuracy, treatment planning, and patient outcomes.

Recent advancements in visual foundation models, such as the Segment Anything Model (SAM) [18], have demonstrated exceptional promise in imparting model generalization for medical image analysis [19]. SAM has exhibited the aptitude for few-shot generalization and shape sensitivity that could potentially address challenges in automatic ovarian tumor segmentation.

In this work, we present MBA-Net, a novel SAM-driven bidirectional aggregation network that effectively adapts and extends SAM's powerful segmentation capabilities to the ovarian tumor domain. The core innovation lies in synergistically composing SAM with a specialized CNN encoder branch through bidirectional feature aggregation. Specifically, our proposed network employs a hybrid encoder architecture consisting of two parallel branches: 1) a prior branch that inherits the SAM encoder to capture robust segmentation priors, and 2) a domain branch composed of multiple residual blocks, specifically designed for extracting domain-specific features. The shallow features from the prior branch are progressively aggregated into the domain branch via a cascade of robust feature injection networks (RFIN). RFIN enables the domain branch to incorporate the robust segmentation priors captured by the prior branch, guiding the extraction of domain-specific features. Conversely, the deep features from the domain

branch are reciprocally integrated into the prior branch through several domain knowledge integration networks (DKIN). DKIN allows the prior branch to leverage the domain-specific knowledge extracted by the domain branch, enhancing its ability to adapt to the unique characteristics of ovarian tumors.
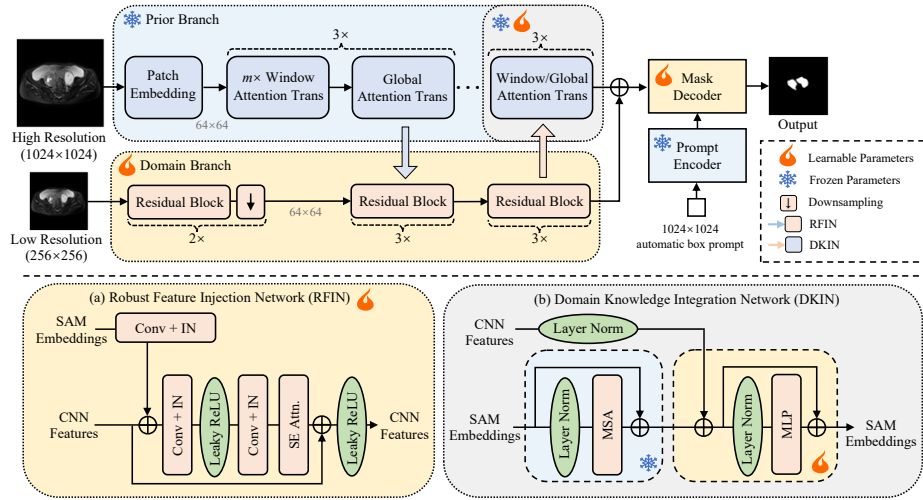

Fig. 1. The architecture of MBA-Net. It has two parallel branches: (1) the prior branch captures robust features, and (2) the domain branch extracts domain-specific features. The bidirectional feature aggregation between the two branches is achieved through (a) the robust feature injection network (RFIN) that injects SAM embeddings into the domain branch and (b) the domain knowledge integration network (DKIN) that integrates CNN features into the prior branch. The prompt encoder and mask decoder are inherited from SAM for mask prediction.

We extensively evaluate MBA-Net on two challenging datasets: a public dataset of ovarian tumor ultrasound images and an in-house dataset of ovarian tumor MRI scans. Comprehensive experiments and comparisons against state-of-the-art methods demonstrate MBA-Net's superior segmentation accuracy, robustness to tumor heterogeneity, and strong generalization across imaging modalities and clinical centers.

2 Methodology

2.1 Overview of MBA-Net

MBA-Net is a novel architecture that synergistically integrates the powerful general segmentation capabilities of the SAM with a domain-specific branch tailored for ovarian tumor segmentation. As depicted in Fig. 1, the encoder of MBA-Net comprises two parallel branches: the prior branch and the domain

branch. The prior branch leverages the pre-trained image encoder from SAM. This branch operates on high-resolution images to capture global contextual information. In contrast, the domain branch is specifically designed to extract domain-specific features from lower-resolution images. It consists of a series of residual blocks with squeeze-and-excitation modules [20], enabling the learning of discriminative tumor representations.

The decoder of MBA-Net inherits the mask decoder from SAM. The fused features from the encoder are passed through the decoder to obtain the final segmentation output. Similarly, MBA-Net also inherits SAM's prompt encoder. We feed a box prompt the same size as the image into the prompt encoder for automatic segmentation.

During training, the prior and domain branches process their respective input images to extract image embeddings. These embeddings are then fused through the bidirectional aggregation, enabling the model to learn a unified representation that combines the strengths of both branches. The decoder takes the fused embeddings as input and generates the segmentation masks.

2.2   The Domain Branch

The domain branch is designed to extract domain-specific features from the input image x_c with a resolution of 256×256. This branch consists of 8 layers, each implemented as a residual block with channel attention.

Let f_c^i denote the output feature map of the i-th layer in the domain branch. The first two layers of the domain branch are responsible for downsampling the input image x_c from a resolution of 256×256 to 64×64. Starting from the third layer, the domain branch receives embeddings from the prior branch at specific locations. The prior branch, which operates on a higher-resolution image x_s, has a total of 4m layers arranged in a transformer architecture (m=8 in our work). The layers at positions {m, 2m, 3m} of the prior branch are global attention layers, enabling the capture of global contextual information. These global attention layers are connected to the {3, 4, 5} layers of the domain branch, respectively.

The feature maps are transmitted back to the prior branch in the later layers of the domain branch, specifically layers {6, 7, 8}. These layers aim to provide domain-specific features to the prior branch, enabling it to adapt its representations to the characteristics of ovarian tumors. The feature transmission is achieved through residual connections, where the output feature maps of the domain branch are added to the embeddings of the prior branch at layers {4m-2, 4m-1, 4m}. Finally, the output of the domain branch and the prior branch are combined through element-wise addition.

2.3   Bidirectional Feature Aggregation

Our architecture is fundamentally anchored in its bidirectional feature aggregation mechanism, a critical structure enabling seamless information flow between

the prior and domain branches. This mechanism fosters a dynamic interchange of features and effectively enhances the network's ability.

Robust Feature Injection Network: As illustrated in Fig. 1a, we devise lightweight residual connections to enable the robust features of the prior branch to be injected into the shallow layers of the domain branch.

We denote the output embedding generated by the i-th layer of the prior branch as $f_i^s \in \mathbb{R}^{N \times C}$, where N is the number of tokens. Likewise, the j-th residual block generates a counterpart feature map $f_j^c \in \mathbb{R}^{H \times W \times C_c}$, where H, W, and C_c represent its height, width and number of channels respectively.

To fuse robust features into the domain branch, a convolution layer is first applied to project f_i^s into an aligned space:

$$\tilde{f}_i^s = \sigma(IN(W_{af} * R(f_i^s) + b_{af})) \tag{1}$$

where $W_{af} \in \mathbb{R}^{C \times C_c \times k \times k}$ and $b_{af} \in \mathbb{R}^{C_c}$ represent the weight and bias of the convolution layer transforming f_i^s from C to C_c channels. R represents reshaping two-dimensional embeddings into three-dimensional features. IN and σ denote instance normalization and Leaky ReLU activation. The projected embedding $\tilde{f}_i^s$ is then combined with f_j^c using residual addition:

$$f_j^{c\prime} = f_j^c + \tilde{f}_i^s, \quad \forall j \in \mathcal{J} \tag{2}$$

where $\mathcal{J}$ denotes the index set $\{3, 4, 5\}$ for low-level residual blocks.

Domain Knowledge Integration Network: Our proposed DKIN is shown in Fig. 1b. We denote the output feature map of the j-th residual block as $f_j^c \in \mathbb{R}^{H \times W \times C}$ and the output embedding generated by the i-th transformer layer as $f_i^s \in \mathbb{R}^{N \times C}$.

To enable domain-specific features to enhance transformer layers in the prior branch, we apply residual addition to combine f_j^c with f_i^s:

$$f_i^{s\prime} = \mathrm{MSA}(\mathrm{LN}(f_i^s)) + \mathrm{LN}(R(f_j^c)) + f_i^s, \quad \forall i \in \mathcal{I} \tag{3}$$

where LN and MSA denote layer normalization and multi-head self attention module. R represents reshaping three-dimensional features into two-dimensional embeddings. $\mathcal{I}$ denotes the index set $\{4m\text{-}2, 4m\text{-}1, 4m\}$ for high-level transformer layers. This top-down dissemination enables advanced domain knowledge to directly guide deeper transformer layers in a tailored manner.

3 Experiments and Results

3.1 Datasets and Implementation Details

To validate the effectiveness of our proposed MBA-Net, we evaluate our method on a public multi-modal ovarian tumor ultrasound segmentation dataset [17]

Table 1. Quantitative comparison of our proposed MBA-Net against other state-of-the-art approaches on the multi-modality ovarian tumor ultrasound dataset. The best results are highlighted in bold.

Methods	Chocolate cyst	Serous cystadenoma	Teratoma	Theca cell tumor	Simple cyst	Mucinous cystadenoma	High grade serous
Dataset: Multi-modality ovarian tumor ultrasound [17]							
U-Net	86.48±15.88	92.40±9.83	83.45±16.99	89.32±6.15	89.67±13.83	89.10±11.68	83.16±12.54
TransFuse	82.40±21.09	90.52±10.69	80.15±15.30	89.19±7.07	90.64±12.26	90.41±8.52	82.34±11.02
TransUNet	82.38±19.67	91.06±11.53	79.33±16.58	89.24±8.16	87.18±12.33	89.32±10.34	81.31±11.77
UTNet	88.54±13.80	**93.78±6.42**	86.59±13.58	91.90±4.67	91.02±13.97	90.32±6.97	84.22±11.44
MBA-Net	**90.39±11.01**†	92.98±5.58~	**89.34±7.75**‡	**92.31±5.11**~	**92.22±11.92**†	**90.67±6.70**~	**88.75±7.16**‡
Cross-modality performance							
U-Net	69.44±27.50	83.75±13.10	60.43±29.05	77.75±19.59	75.47±30.43	85.98±13.60	48.85±32.38
TransFuse	76.74±17.96	83.52±13.77	62.57±28.01	76.30±18.81	75.42±31.42	88.61±7.50	47.55±33.59
TransUNet	76.13±19.43	85.74±13.92	60.75±28.23	77.03±18.53	78.84±26.21	87.03±11.84	45.88±33.38
UTNet	75.55±14.82	81.53±14.18	61.28±30.26	73.77±19.07	77.51±25.54	81.56±7.26	50.34±31.61
MBA-Net	**78.89±13.36**‡	**86.51±12.68**‡	**68.86±26.01**‡	**80.64±12.67**‡	**82.98±21.18**‡	**90.81±5.85**†	**70.60±22.79**‡

~:p-value>0.05, †:p-value>$1e^{-3}$, ‡:p-value<$1e^{-3}$, in comparison to the second-best or best results.

and an in-house multi-site ovarian tumor MRI dataset. We employ stratified sampling to maintain a consistent distribution of tumor types, using a 6:1:3 split for training, validation, and testing for both datasets. Additionally, to assess the robustness of the model, we perform cross-domain and cross-modality testing using all available images.

After excluding samples of normal ovaries, the ultrasound dataset consists of seven different tumor types: chocolate cyst, serous cystadenoma, teratoma, theca cell tumor, simple cyst, mucinous cystadenoma, and high-grade serous carcinoma. We utilize the ultrasound images for training, evaluating the model's performance and assessing its robustness by contrast-enhanced ultrasonography images.

The MRI dataset comprises 493 patients from five medical centers, encompassing five distinct tumor types: serous cystadenocarcinoma, mucinous cystadenocarcinoma, serous cystadenoma, mucinous cystadenoma, and clear cell ovarian carcinoma. We employ data from one center for training and evaluating the model's performance while leveraging the other four centers to assess its robustness. Detailed information about both datasets is provided in Appendix.

We compare MBA-Net with several state-of-the-art medical image segmentation networks, including U-Net [21], TransFuse [22], TransUNet [23], and UT-Net [24]. It is important to note that direct comparison between our fully automatic MBA-Net and interactive segmentation models like SAM and Med-SAM [25] is not appropriate. These models rely on user-provided prompts, which gives them an inherent advantage in localization but limits their applicability in fully automated workflows. We use Dice score as the evaluation metric and Wilcoxon signed-rank tests to assess statistical significance between our MBA-Net and other methods. We apply consistent data augmentation techniques to ensure a fair comparison among different methods, including random intensity scaling, random intensity shifting, and random flipping. The training and test-

Table 2. Quantitative comparison of our proposed MBA-Net against other state-of-the-art approaches on the multi-site ovarian tumor MRI dataset. The best results are highlighted in bold.

Methods	Serous cystadenocarcinoma	Mucinous cystadenocarcinoma	Serous cystadenoma	Mucinous cystadenoma	Clear cell ovarian carcinoma
Dataset: Multi-site ovarian tumor MRI					
U-Net	78.33±21.94	82.09±20.75	82.24±15.22	90.02±9.80	89.47±10.10
TransFuse	81.17±19.27	81.48±17.97	79.44±18.06	92.94±7.59	89.88±12.23
TransUNet	80.91±18.41	83.56±15.72	85.16±12.53	91.36±8.62	86.92±10.08
UTNet	79.69±21.05	83.82±16.67	**86.69±14.40**	91.45±6.83	91.61±8.94
MBA-Net	**83.51±17.14‡**	**86.75±12.18‡**	86.42±11.96$^{\sim}$	**93.79±6.55†**	**94.73±5.50‡**
Cross-site performance					
U-Net	74.05±23.88	69.75±23.68	71.14±25.39	90.33±11.90	81.73±19.96
TransFuse	73.11±22.06	70.79±21.41	75.72±20.26	88.49±14.74	80.29±21.33
TransUNet	72.99±20.11	75.42±17.05	78.72±18.88	89.54±15.44	85.84±14.62
UTNet	75.79±22.41	77.92±18.13	74.40±21.56	90.77±13.82	84.47±13.77
MBA-Net	**80.35±19.29‡**	**83.47±16.26‡**	**83.18±14.97‡**	**91.19±11.76$^{\sim}$**	**89.11±11.92‡**

$\sim$:p-value>0.05, $\dagger$:p-value>$1e^{-3}$, $\ddagger$:p-value<$1e^{-3}$, in comparison to the second-best or best results.

ing processes are conducted on an RTX 4090 GPU with 24GB of memory. Each model is trained for 50 epochs. During the training phase, we employ the SGD optimizer, and the initial learning rate is 3e-4.

3.2   Results

We evaluate the performance of MBA-Net and compare it with state-of-the-art segmentation methods on the multi-modality ovarian tumor ultrasound dataset. The results are summarized in Table 1. MBA-Net achieves the highest average Dice score, outperforming all other methods. To assess the robustness of MBA-Net, we perform cross-modality evaluation by testing the model on contrast-enhanced ultrasonography images. As shown in Table 1, MBA-Net exhibits superior generalization capability, surpassing all other methods by a significant margin. Remarkably, MBA-Net has substantially improved the segmentation of teratoma and high-grade serous carcinoma.

We further validate the performance of MBA-Net on the multi-site ovarian tumor MRI dataset. The results are presented in Table 2. MBA-Net consistently outperforms other methods across most tumor types. It also demonstrates exceptional generalization ability in the cross-site setting. The superior performance of our proposed method in these scenarios demonstrates its potential to enhance standardization of ovarian tumor assessment across different imaging protocols and healthcare institutions, potentially leading to more consistent diagnosis and treatment planning.

Fig. 2 presents the visual comparison of the segmentation results obtained by different methods on representative examples from both two datasets, illustrating the superior performance of MBA-Net in accurately delineating ovarian tumors with diverse morphological characteristics and imaging modalities. Notably, MBA-Net shows improved performance in capturing fine details and

complex boundaries, particularly in cases where other methods struggle with heterogeneous tumor appearances or ambiguous edges.
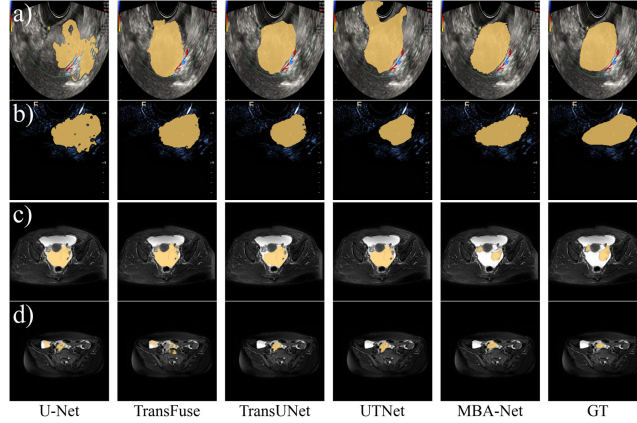

Fig. 2. Visual comparison of segmentation results obtained by different methods on representative examples from (a-b) the multi-modality ovarian tumor ultrasound dataset and (c-d) the multi-site ovarian tumor MRI dataset. GT: ground truth.

Ablation Studies To investigate the effectiveness of the proposed bidirectional feature aggregation mechanism, we conduct ablation studies on the ovarian tumor ultrasound dataset. We evaluate the impact of varying the number of RFIN and DKIN modules on the segmentation performance. The results are summarized in Table 3. The best performance is achieved when both RFIN and DKIN modules are utilized, with the default setting of 3 RFIN and 3 DKIN modules yielding an average Dice score of 90.75%. This result validates the effectiveness of the bidirectional feature aggregation mechanism in leveraging the complementary strengths of the prior branch and the domain branch.

4   Conclusion

In this paper, we present MBA-Net, a novel architecture for ovarian tumor segmentation that integrates the powerful segmentation capabilities of the SAM with domain-specific knowledge through bidirectional feature aggregation. The proposed network employs a hybrid encoder architecture, where the encoder consists of a prior branch, which inherits the SAM encoder to capture robust segmentation priors, and a domain branch, which is specifically designed to extract domain-specific features. The robust feature injection network (RFIN) and the domain knowledge integration network (DKIN) facilitate the bidirectional flow of information between the two branches, enabling MBA-Net to leverage

the complementary strengths of both branches. Extensive experiments on multi-modality ovarian tumor ultrasound and multi-site ovarian tumor MRI datasets demonstrate the superiority of MBA-Net in terms of segmentation accuracy, robustness to tumor heterogeneity, and generalization capability across different imaging modalities and clinical sites.

Table 3. Ablation studies of the different feature fusion strategies.

RFIN	DKIN	Avg. Dice (%)	RFIN	DKIN	Avg. Dice (%)	RFIN	DKIN	Avg. Dice (%)
0	1	85.84	0	3	86.15	0	6	86.58
1	1	88.69	1	3	88.43	1	6	89.06
2	1	88.01	2	3	90.19	2	6	89.82
3	1	88.25	3	3	**90.75**	3	6	90.47

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. 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.
2. Lena Sophie Kiefer, Jana Fabian, Roberto Lorbeer, Jürgen Machann, Corinna Storz, Mareen Sarah Kraus, Elke Wintermeyer, Christopher Schlett, Frank Roemer, Konstantin Nikolaou, et al. Inter-and intra-observer variability of an anatomical landmark-based, manual segmentation method by mri for the assessment of skeletal muscle fat content and area in subjects from the general population. *The British Journal of Radiology*, 91(1089):20180019, 2018.
3. Arne Schmidt, Pablo Morales-Álvarez, and Rafael Molina. Probabilistic modeling of inter-and intra-observer variability in medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21097–21106, 2023.
4. Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H Schwartz, and Hugo JWL Aerts. Artificial intelligence in radiology. *Nature Reviews Cancer*, 18(8):500–510, 2018.
5. Jeffrey D Rudie, Andreas M Rauschecker, R Nick Bryan, Christos Davatzikos, and Suyash Mohan. Emerging applications of artificial intelligence in neuro-oncology. *Radiology*, 290(3):607–618, 2019.

6. Jianfei Liu, Shijun Wang, Marius George Linguraru, Jianhua Yao, and Ronald M Summers. Tumor sensitive matching flow: A variational method to detecting and segmenting perihepatic and perisplenic ovarian cancer metastases on contrast-enhanced abdominal ct. *Medical image analysis*, 18(5):725–739, 2014.
7. Travis W Sawyer, Photini FS Rice, David M Sawyer, Jennifer W Koevary, and Jennifer K Barton. Evaluation of segmentation algorithms for optical coherence tomography images of ovarian tissue. *Journal of Medical Imaging*, 6(1):014002–014002, 2019.
8. Zihang Liu, Chunhui Zhao, Yan Lu, Yu Jiang, and Jingtian Yan. Multi-scale graph learning for ovarian tumor segmentation from ct images. *Neurocomputing*, 512:398–407, 2022.
9. Jiaqi Hu, Zhiming Cui, Xiao Zhang, Jiadong Zhang, Yuyan Ge, Honghe Zhang, Yan Lu, and Dinggang Shen. Uncertainty-aware refinement framework for ovarian tumor segmentation in cect volume. *Medical Physics*, 2023.
10. Lijiang Chen, Changkun Qiao, Meijing Wu, Linghan Cai, Cong Yin, Mukun Yang, Xiubo Sang, and Wenpei Bai. Improving the segmentation accuracy of ovarian-tumor ultrasound images using image inpainting. *Bioengineering*, 10(2):184, 2023.
11. Dingdu Hu, Junming Jian, Yongai Li, and Xin Gao. Deep learning-based segmentation of epithelial ovarian cancer on t2-weighted magnetic resonance images. *Quantitative Imaging in Medicine and Surgery*, 13(3):1464, 2023.
12. Yifan Gao, Yin Dai, Fayu Liu, Weibing Chen, and Lifu Shi. An anatomy-aware framework for automatic segmentation of parotid tumor from multimodal mri. *Computers in Biology and Medicine*, 161:107000, 2023.
13. S Nougaret, M Tardieu, HA Vargas, C Reinhold, S Vande Perre, N Bonanno, E Sala, and I Thomassin-Naggara. Ovarian cancer: an update on imaging in the era of radiomics. *Diagnostic and interventional imaging*, 100(10):647–655, 2019.
14. Mandi Wang, Jose AU Perucho, Yangling Hu, Moon Hyung Choi, Lujun Han, Esther MF Wong, Grace Ho, Xiaoling Zhang, Philip Ip, and Elaine YP Lee. Computed tomographic radiomics in differentiating histologic subtypes of epithelial ovarian carcinoma. *JAMA Network Open*, 5(12):e2245141–e2245141, 2022.
15. Robert A Soslow. Histologic subtypes of ovarian carcinoma: an overview. *International Journal of Gynecological Pathology*, 27(2):161–174, 2008.
16. Thi-Loan Pham, Van-Hung Le, Thanh-Hai Tran, and Duy Hai Vu. Comprehensive study on semantic segmentation of ovarian tumors from ultrasound images. In *Conference on Information Technology and its Applications*, pages 262–273. Springer, 2023.
17. Qi Zhao, Shuchang Lyu, Wenpei Bai, Linghan Cai, Binghao Liu, Meijing Wu, Xiubo Sang, Min Yang, and Lijiang Chen. A multi-modality ovarian tumor ultrasound image dataset for unsupervised cross-domain semantic segmentation. *arXiv preprint arXiv:2207.06799*, 2022.
18. Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollar, and Ross Girshick. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4015–4026, October 2023.
19. Yifan Gao, Wei Xia, Dingdu Hu, and Xin Gao. Desam: Decoupling segment anything model for generalizable medical image segmentation. *arXiv preprint arXiv:2306.00499*, 2023.
20. Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

21. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

22. Yundong Zhang, Huiye Liu, and Qiang Hu. Transfuse: Fusing transformers and cnns for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 14–24. Springer, 2021.

23. Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

24. Yunhe Gao, Mu Zhou, and Dimitris N Metaxas. Utnet: a hybrid transformer architecture for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, pages 61–71. Springer, 2021.

25. Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.