Application of AI large models for risk reduction of gynecological tumor diagnosis and treatment

Fa Qiang Qian[a], Kunze Yang[b], Ming Yang[a*] ,Si Juan Chen[a*] , Kun Shi[c], Chenxi Zhang[a],Jianhua Chen[a]

[a] Shenzhen Key Leboratory of Nuclear and Radiation Safety,Institute for Advanced Study in Nuclear Energy & Safety, College of Physics and Optoelectronic Engineering,
Shenzhen University, Shenzhen 518060, China
[b] Pennsylvania State University, Pennsylvania, America
[c] Guangzhou Women and Children's Medical Center, Guangzhou, China
[*] **Corresponding author:**mingyang@szu.edu.cn;chensijuan@szu.edu.cn)

Abstract: Gynecological tumors pose a significant health risk to women, with malignant tumors threatening their lives. Accurate screening relies on the interpretation of medical images by doctors, which can often be blurry, noisy, or have low contrast, introducing risks into medical diagnoses. Even experienced physicians may not always make entirely accurate judgments. In the diagnosis and treatment of gynecological tumors, misdiagnosis can severely impact patient care and mental health, as well as present major challenges in hospital management.

Traditional medical image feature extraction requires manual operations and often yields unsatisfactory results. With the development of deep learning, methods such as CNNs, including Unet, have shown excellent performance in medical imaging tasks. The large-scale vision model "Segmentation of Everything" (SAM), released by MetaAI in 2023, caused a sensation in the field of image segmentation. SAM consists of an image encoder, a prompt encoder, and a mask decoder. MedSAM incorporated medical domain knowledge into this framework, building a general model for medical image segmentation. Based on this, we aim to apply it to lesion or organ recognition in gynecology to reduce the risk of misdiagnosis in gynecological tumors. We propose the USFP (UnetSam for Pelvic segmentation) model, which freezes the prompt encoder and adds a Unet module. Features extracted by Unet are then fed into the SAM model, followed by a residual connection with the image encoder. This model was fine-tuned and validated on the public BTCV_Cervix dataset for pelvic segmentation, further improving the Dice coefficient for pelvic organ segmentation, achieving scores of 94.4%, 97.77%, 93.21%, and 91.7% respectively for different organs. These results surpass those of fine-tuned MedSAM and SAM. Additionally, visualization shows smoother edges, effectively alleviating the issue of jagged borders, making the analysis more robust, which is significant for reducing the risk of misdiagnosis in the diagnosis and treatment of gynecological tumors.

Keywords: Gynecological tumors , Deep learning, artificial intelligence,Medical risks

1. INTRODUCTION

Gynecological tumors, as one of the major diseases threatening women's health, affect not only the physical well-being of patients but also profoundly impact their quality of life and psychological state. Statistics show that the incidence of gynecological malignancies is rising annually on a global scale, presenting a significant challenge to public health. Among these, cancers such as cervical and ovarian cancer are particularly prevalent, and the accuracy of pre-surgical diagnoses directly correlates with treatment outcomes and patient prognoses. However, within current diagnostic procedures, despite the critical visual information provided by medical imaging technologies like MRI and CT, the complexity of these images and potential interfering factors (such as blurriness, noise pollution, and insufficient tissue contrast) often lead to diagnostic difficulties. Even highly experienced clinicians may encounter misdiagnoses or missed diagnoses due to subjective judgment variations. A recent case of misdiagnosis on May 11, 2024, at Hong Kong Poai Hospital, where a patient was incorrectly diagnosed with endometrial cancer, sparked considerable online attention. This erroneous pre-surgical judgment led to the discovery post-surgery that there were no cancerous lesions, causing irreversible damage to the patient's physical and mental health, as well as to their family members. Therefore, enhancing the accuracy of pre-surgical diagnoses is of paramount importance.

In recent years, the breakthrough progress of artificial intelligence (AI) technology, especially deep learning algorithms, in the field of medical image analysis has brought new hope for improving diagnostic accuracy. Convolutional neural networks (CNNs) such as U-net have shown significant advantages in medical image segmentation tasks due to their powerful feature extraction capabilities, providing strong support for accurately identifying lesion boundaries and assisting clinical decision-making. In 2023, MetaAI released a

large-scale visual model "Segment Anything Model" (SAM), which effectively utilizes the global context information of the image by introducing a self-attention mechanism, greatly improving the accuracy and generalization of image segmentation. However, when the SAM model is directly applied to medical images without specific field adjustments, especially the diagnosis of gynecological tumors, its performance is limited and fails to fully meet clinical needs. On this basis, MedSAM has been fine-tuned with a large number of medical images, giving it certain medical image knowledge, but it still has limitations in specific task segmentation.

In view of this, this study focuses on exploring how to optimize and apply large models, especially by combining medical expertise to transform and fine-tune the MedSAM model to develop AI-assisted tools that are more suitable for the diagnosis and treatment of gynecological tumors. We innovatively introduced the USFP (UnetSAM for Pelvic segmentation) model and made targeted improvements to it to address the inherent challenges in gynecological pelvic MRI image analysis. Specifically, we designed a new fusion model by freezing the prompt encoder in SAM, integrating the U-net structure for feature enhancement, and then performing a residual connection with the image encoder. This model can not only effectively extract and fuse deep and shallow features, but also significantly improve the recognition accuracy of pelvic organs and potential tumor areas through fine-tuning on a dedicated gynecological pelvic segmentation dataset BTCV_Cervix, surpassing the performance of the baseline model.

This study not only verifies the potential of large model transformation in improving the accuracy of gynecological tumor diagnosis and treatment, but also provides an empirical basis and methodological guidance for the future integration of AI technology into gynecological tumor clinical practice more widely and deeply. Our research has four main advantages:

- **Enhanced spatial understanding:** U-Net's pre-processing enhances the model's sensitivity to details and edges, especially in the recognition of complex anatomical structures, which is crucial for the accurate segmentation of cervical cancer.
- **Global and local feature fusion:** Combining the strategies of U-Net and Transformer, the seamless integration of local details and global context is achieved, which improves the accuracy and robustness of segmentation.
- **Improve model generalization ability:** Through such multi-scale feature fusion, the model can better generalize to unseen data, especially when facing the common variability in medical imaging data.
- **Have higher segmentation indicators:** Our experiments show that our proposed method has better segmentation indicators, which greatly reduces the risk of misdiagnosis in gynecological imaging diagnosis.

The following chapters will elaborate on our methodology, experimental design, result analysis, and discussion of research findings in detail, in order to provide scientific basis and technical support for reducing the risk of misdiagnosis in gynecological tumor diagnosis and treatment.

2. Related Work

2.1 Application of medical imaging in the diagnosis of gynecological tumors

Medical imaging technologies, especially MRI and CT scans, have become indispensable tools in the diagnosis of gynecological tumors. These technologies provide detailed anatomical information, assisting doctors in identifying the location, size, and relationship of tumors with surrounding tissues. However, the radiological manifestations of gynecological tumors are diverse and complex, often being easily confused with benign conditions, which increases the difficulty of diagnosis. For example, endometriosis and uterine fibroids can appear similar to certain malignant tumors on MRI images, increasing the likelihood of misdiagnosis. Therefore, enhancing the precision of image analysis is crucial for reducing misdiagnoses and missed diagnoses. Our approach aims to further analyze medical images using large AI models to decrease the rate of misdiagnosis.

2.2 Limitations of traditional gynecological tumor diagnosis methods

Traditionally, the diagnosis of gynecological tumors primarily relies on the intuitive analysis of medical images by physicians, combined with clinical symptoms, laboratory tests, and pathological examination results. This process is highly dependent on the physician's experience and subjective judgment, making it susceptible to individual variations. Moreover, the complexity of the female pelvic anatomy, coupled with inherent technical limitations of the images themselves—such as inadequate resolution and poor contrast—makes accurately distinguishing between normal tissue and lesions particularly challenging. The results of our method

can serve as a supplement to physician analysis, providing a clear segmentation to aid in the differentiation of lesions.

2.3 Progress of AI technology in gynecological tumor diagnosis

In recent years, with the rapid development of artificial intelligence, especially deep learning technology, its application in the field of medical image analysis has become increasingly widespread. Many studies have shown that models based on deep learning, such as convolutional neural networks (CNNs), can effectively improve the detection and classification accuracy of gynecological tumors. For example, some studies have used the Unet architecture to segment gynecological MRI images, improving the localization accuracy of tumors such as cervical cancer and ovarian cancer. Then in 2017, the Transformer architecture came out, originally designed for natural language processing tasks, and has demonstrated its strong potential in processing image tasks. In particular, in medical image segmentation, Vision Transformer (ViT) demonstrated how to compete with CNN by segmenting images into sequences. In 2021, Chen et al. introduced TransUNet, a hybrid architecture that combines CNN and Transformer, which takes advantage of the powerful spatial localization ability of U-Net and the long-distance dependency modeling ability of Transformer. In addition, some studies have tried to combine multimodal imaging data, such as PET-CT, to further improve the specificity and sensitivity of diagnosis.

2.4 The rise of large pre-trained models and fine-tuning

Recently, with the success of large language models such as BERT pre-trained and fine-tuned in the field of natural language processing, the emergence of large pre-trained models such as "Segmentation of Everything" (SAM) has also brought revolutionary changes to the field of image processing. These models learn rich visual representations by pre-training on large-scale non-medical data, and then show great potential in specific medical image analysis tasks through domain adaptation and fine-tuning strategies. However, when directly applying these models to specialized fields such as gynecological oncology, it is often necessary to overcome domain gaps and ensure that the models can understand and handle the unique complexity and nuances of medical images. MedSAM freezes the image encoder and focuses on adjusting SAM's mask decoder to train on medical images at an acceptable cost. SAMed uses a low-rank (LoRA) strategy to tune the image encoder to adjust SAM at a lower computational cost, making it more suitable for medical image segmentation. MSA uses two downReLU-upAdapters on each transformation layer of the ViT image encoder to introduce task-specific information.

Despite the significant technological advances mentioned above, there are still challenges in effectively applying large AI models to reduce the risk of misdiagnosis in the diagnosis and treatment of gynecological tumors. How to effectively optimize the characteristics of gynecological pelvic images while maintaining the generalization ability of the model to achieve high-precision lesion identification and segmentation is an important gap in current research. In addition, how to effectively integrate medical expertise and AI technology to improve the diagnostic efficiency of the model is also an issue that needs to be addressed.

3. Methodology

3.1. Data Processing

Gynecological MRI image analysis faces many challenges, including but not limited to complex pelvic anatomy, low soft tissue contrast in images, and variability introduced by different scanning postures and patient preparation states. These factors work together to increase the difficulty of image analysis and lesion identification, especially for automated computer vision algorithms.

To address the above challenges, this study selected the BTCV_cervix dataset as the basic resource. This dataset is designed for medical imaging research related to cervical cancer. It contains 30 training data with detailed annotations and 20 unlabeled test data, aiming to promote the study of cervical cancer treatment planning. The images in the dataset are stored in .nii.gz format, which is a commonly used medical imaging format suitable for compressed storage of three-dimensional image data.

In the dataset, each patient's MRI image is labeled with key structures such as bladder, uterus, rectum, and small intestine. These labels are crucial for understanding the relationship between tumors and surrounding tissues. It is worth noting that the scans in the dataset follow a specific preparation procedure, requiring patients to drink water to fill their bladders before scanning, which helps to enhance image contrast and improve diagnostic accuracy. In addition, although all patients had markers implanted around their cervix, these markers

were only used for research purposes and were not directly involved in image analysis. Table 1 shows the information of this dataset.

Table 1. BTCV_Cervix Dataset

Anatomical structure	**Bladder**	**Uterus**	**Rectum**	**Small intestine**
Number of occurrences	30	30	30	28
Percentage of occurrences	100%	100%	100%	93%
Minimum volume (cm^3)	64	62	29	209
Median volume (cm^3)	287	197	87	559
Maximum volume (cm^3)	665	503	227	1676
Number of structure slices	1015	1305	1145	1390

Considering the requirements of model input and the actual situation of the dataset, the following preprocessing steps were taken, and Figure 1 shows the data preprocessing process:

- From 3D to 2D: First, 2D slices were extracted from the 3D volume images in .nii.gz format to adapt to subsequent model processing. This conversion allows us to use a large number of existing 2D image processing techniques.
- Resizing: The extracted 2D slices were resized to a size compatible with the Segment-Anything (SAM) model, that is, 3 channels and 1024×1024 pixels. For images whose original size is smaller than the model input requirement, the padding technique is used to fill the edges to maintain the integrity of the spatial information.
- Tip box generation: According to the four anatomical structure labels provided, a tip box (bounding box) is automatically generated for each slice. The specific operation is to take the upper left and lower right corner coordinates of each label shape. These boxes are used to guide the encoder of the SAM model to focus on key areas, thereby improving the accuracy of segmentation and recognition.
- Data partitioning: From the 30 labeled data, 80% of the slices were randomly selected for model training, and the remaining 20% were used for validation to ensure the generalization ability of the model. This process generated a total of 5527 2D slices, of which 3009 lacked clear label information, reflecting the uncertainty often found in actual clinical data.

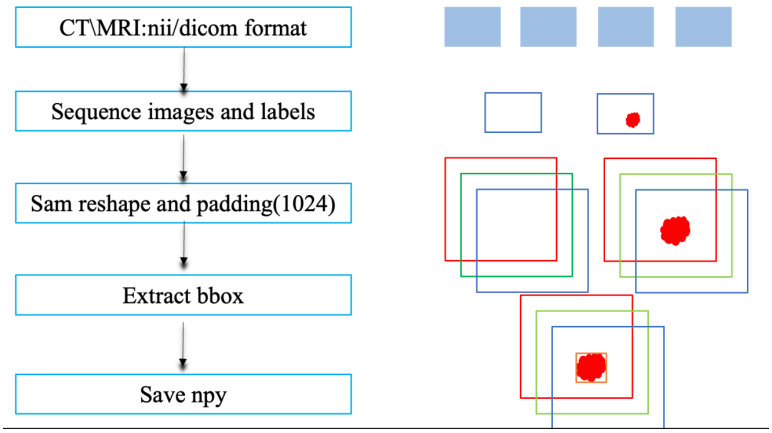

Figure 1. Data Preprocessing

Through the above-mentioned meticulous data processing process, not only the difficulties in analyzing the original MRI images were solved, but also a solid foundation was laid for the subsequent use of the SAM model for efficient and accurate risk assessment of gynecological diseases.

3.2. Fine-tuning strategies for SAM and MedSAM models

Segment-Anything (SAM) is an innovative instance segmentation framework, which is based on flexible and efficient segmentation capabilities, and is particularly suitable for handling challenging medical imaging tasks. It is modified based on transformer, which is mainly composed of attention mechanism and MLP. The formula of attention mechanism is as follows:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

The core architecture of SAM consists of three parts: image encoder, prompt encoder, and mask decoder. SAM uses the pre-trained Vision Transformer (ViT), specifically the ViT-H/16 variant, which has a 14×14 window attention mechanism and four layers of equally spaced global attention blocks, which can efficiently process high-resolution input images. The encoder is based on MAE pre-training, which ensures strong feature extraction capabilities and outputs image embeddings that are 16 times downsampled from the input image. The prompt encoder converts the manually input points, boxes, and masks into sparse and dense encodings, while the mask decoder performs four operations per layer, including self-attention of tokens, cross-attention from tokens to image embeddings, point-based multi-layer perceptron (MLP) token updates, and reverse cross-attention from image embeddings to tokens, which allows the image embeddings to incorporate prompt information. The mask decoder consists of two layers, using residual connections, layer normalization, and dropout (0.1) during training to enhance stability and generalization. In all attention layers, positional encodings are added to the image embeddings to ensure the model's sensitivity to geometric information. At the SAMe time, the original prompt token and its position encoding are re-added to the updated token each time it participates in the attention operation, strengthening the model's dual consideration of the prompt position and type dependence.

Given that the prompt encoder in the SAM model has been fully optimized in the pre-training stage, the training of SAM and MedSAM maintains the SAM targeted strategy as MedSAM during the fine-tuning stage. Specifically, the prompt encoder remains frozen, and the image encoder and mask decoder are trained. The advantage of this strategy is that it not only retains the rich visual features learned by the model from large-scale data, but also enables the model to quickly adapt to the segmentation needs of specific medical images, especially the complex anatomical structures covered by the BTCV_cervix dataset. The training process is shown in Figure 2.

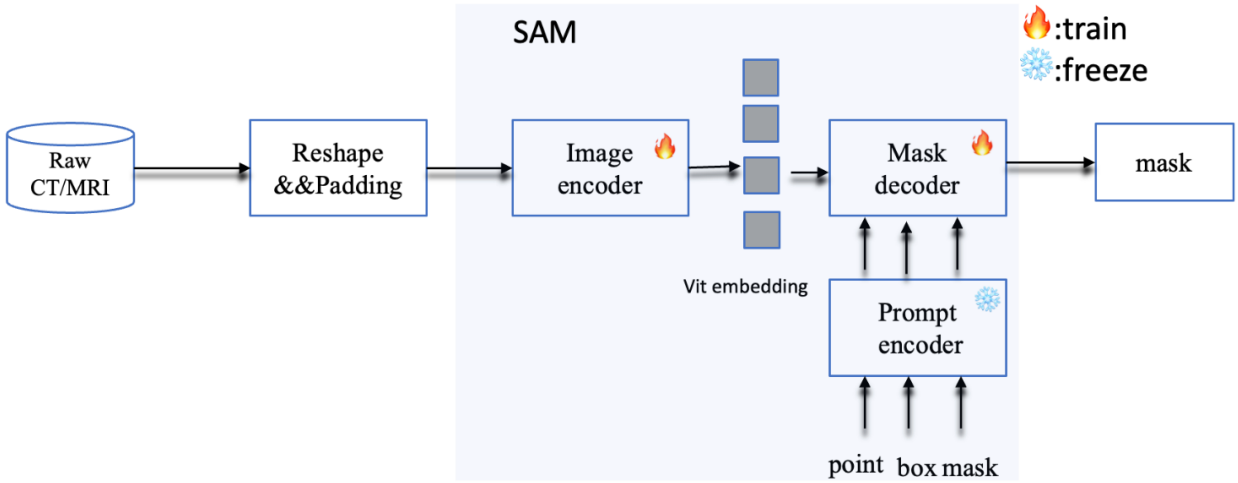

Figure 2. SAM fintune Architecture

The loss function uses DiceLoss and BCEloss. Diceloss is a commonly used loss function in medical image segmentation tasks. It is derived from the similarity coefficient in biology and is particularly suitable for evaluating the overlap between two sets (predicted segmentation areas and true segmentation areas). Let X and Y represent the predicted mask and the true mask respectively. The formula is as follows

$$DiceLoss = 1 - Dice = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (2)$$

BCELoss measures the distribution consistency between binary values. Let x_i and y_i represent each pixel in the mask respectively. The formula of its loss function is shown in Formula 3

$$BCELoss = -\frac{1}{N}\sum_{i=1}^{N}(x_i\, log(\tilde{y_i}) + (1 - x_i)\log(1 - \tilde{y_i})) \quad (3)$$

3.3. USFP

While exploring the balance between segmentation accuracy and model efficiency, this section proposes an innovative architecture improvement strategy called USFP (UnetSAM for Pelvic segmentation) which aims to combine the spatial information capture advantages of the classic U-Net model with the powerful feature expression capabilities of Transformer. This improvement is not only aimed at MedSAM, hoping to maintain the medical knowledge of the model, but also to learn more information in the representation of images, especially in dealing with the complex details of the BTCV_cervix dataset. The figure and formula are shown below

U-Net preliminary feature extraction: First, the original image is preprocessed by a carefully designed U-Net architecture. This U-Net contains four downsampling layers and corresponding upsampling layers, which are designed to extract rich local features and maintain spatial information. U-Net is known for its excellent boundary preservation and detail capture capabilities, which provides a solid foundation for subsequent semantic understanding.

Feature integration and Transformer compatibility: The output of U-Net then passes through a 1×1 convolution layer, which adjusts the number of feature channels to 256 dimensions to match the embedding dimension required by Transformer. This adjustment not only ensures efficient feature transmission, but also facilitates integration with subsequent Transformer components. Then, the adjusted feature map is upsampled to (64, 64) size to align with the feature map obtained from the image encoder, thereby achieving smooth transition and complementarity of information.

Image encoder enhancement: The U-Net features after the above processing are combined with the output of the ViT image encoder in the original SAM. This combination not only retains the global semantic information of the original image (captured by ViT), but also introduces the fine local features extracted by U-Net, providing a more comprehensive visual understanding basis for subsequent mask prediction.

Unified Mask Decoder input: The fused feature map is fed into the improved mask decoder together with the input from the prompt encoder. The design of the mask decoder remains unchanged, but its input now contains deeper image information, combining the spatial accuracy of U-Net and the global context understanding ability of Transformer.

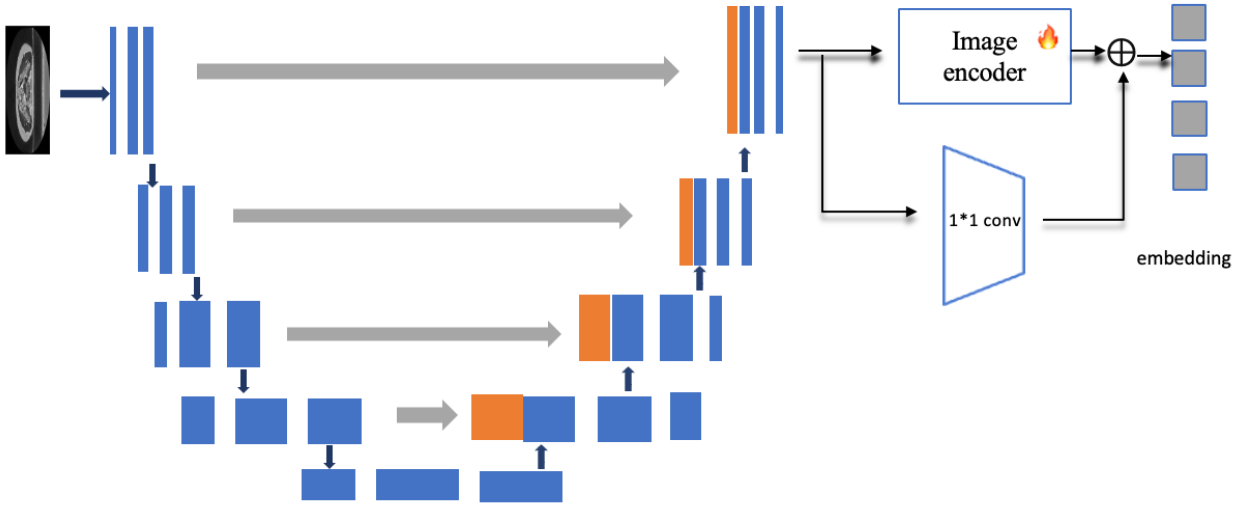

Figure3 USFP embedding Architecture

Assume the original image is I, and the 4-layer U-net is denoted as Unet, then the improved part formula can be expressed as follows

$$embeddings=Unet(I)+image\_embedding(Unet(I)) \quad (4)$$

4. Experiments

4.1 Experimental Design and Implementation

In order to verify the effectiveness of the proposed USFP (Unet-SAM for Pelvic Segmentation) model, we designed a series of detailed experiments, focusing on its performance in four challenging medical image segmentation tasks. These tasks cover the segmentation of different anatomical structures, and each task requires a high degree of accuracy and generalization. The following is the basic framework of the experiment:

Dataset: The BTCV_cervix dataset is used, which contains high-quality pelvic medical images and is suitable for evaluating the segmentation performance of the model in complex scenarios. All images are standardized and divided into training sets, validation sets, and test sets to ensure that the data distribution is balanced and unbiased. Eight copies of each label are used for training and two copies are used for testing, and the number is shown in Table 2

Model Comparison: Three models were selected for comparison, including the original SAM model, the MedSAM model optimized specifically for medical imaging, and our proposed USFP model that combines the advantages of U-Net and Transformer. Since each model performed poorly when directly applied to pelvic data, producing blank or indiscriminate segmentations, we fine-tuned each model separately, keeping the hyperparameter configuration consistent to ensure fairness in the experiment. The training hyperparameters were set as follows: over 50 epochs, batch size of 2, weight decay of 0.001, and a learning rate of 0.0002.

Evaluation Metrics: The Dice Similarity Coefficient (DICE) was used as the primary evaluation metric. The Dice coefficient measures the overlap between predicted labels and ground truth labels. The average Dice score was obtained by averaging across all images in the test set. As shown in Table 3, our model achieved the best results across all four labels on the BTCV_cervix dataset.

Table 2. Data distribution

Type	Bladder	Uterus	Rectum	Small intestine
Train	812	1044	916	1112
Test	203	261	229	278

Table 3. Average Dice Coefficient Experiments Result

Model name	Bladder	Uterus	Rectum	Small intestine
SAM	0.932	0.9057	0.9113	0.8702
MedSAM	0.9409	0.9207	0.923	0.8998
USFP(ours)	0.9441	0.9377	0.9321	0.917

4.2 Visual comparison of results

To visually demonstrate the segmentation performance of each model, a comparative segmentation result image from a typical case is provided below. It can be seen from the figure that the USFP model performs better in maintaining the clarity of structural boundaries, the smoothness of segmentation, and internal consistency, even outperforming the processed gt_mask. Figure 4 shows the comparison results of the bladder and small intestine labels after fine-tuning the three models.

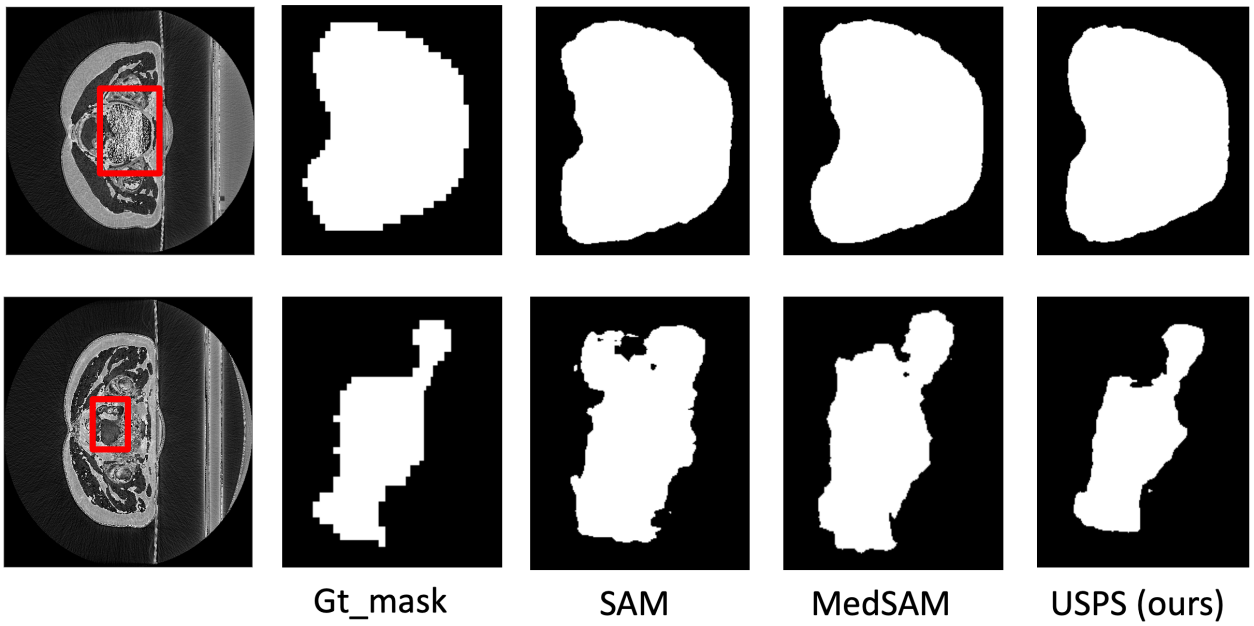

Figure 4 result demo

5 Conclusion

5.1 Conclusion

By comparing the experimental results, the USFP model showed significant performance improvements in all classifications, especially improving the segmentation quality of complex anatomical structures while maintaining high accuracy. This confirms the effectiveness of combining U-Net and Transformer mechanisms, providing a powerful new method for pelvic medical image segmentation. Future work will explore more optimization strategies to further improve the generalization ability and real-time performance of the model.

5.2 Outlook

This study successfully developed and verified the efficiency and accuracy of the USFP (Unet-SAM for Pelvic Segmentation) model in various pelvic anatomical structure segmentation tasks, laying a solid foundation for subsequent clinical applications. However, this is just the first step toward a broader vision of automating medical image analysis. Future exploration will focus on the following key directions:

Lymph node metastasis identification and endometrial cancer localization: Based on the success of the USFP model, the next step will be to expand to a more complex task—using deep learning technology to identify signs of pelvic lymph node metastasis, especially for early diagnosis and staging of endometrial cancer patients. . This new task requires the model to not only accurately segment the lymph node region, but also further determine whether it has metastasized, providing key information for clinical decision-making.

Adaptive and personalized model optimization: Given the physiological differences between patients, developing models that can adaptively learn individual characteristics will be an important direction in the future. This includes, but is not limited to, utilizing large model transfer learning to reduce annotation data requirements and developing dynamically adjustable model parameter strategies to achieve more personalized medical image analysis.

Clinical validation and practical application: Ultimately, the purpose of all technological advancements is to serve clinical practice. Therefore, future work will focus on cooperating with medical institutions to conduct large-scale clinical trials to verify the effectiveness and safety of the model in real-world settings, promote it from the laboratory to the clinic, and truly improve the diagnosis, treatment path and prognosis of patients. .

In summary, although the proposal of the USFP model marks important progress in the field of pelvic image segmentation, the scientific research road ahead is still full of challenges and opportunities. We look forward to making greater contributions to precision medicine and early cancer detection through continuous innovation and optimization, thereby contributing to reducing the risk of misdiagnosis in the diagnosis and treatment of gynecological tumors.

Acknowledgements

This study was supported by Shenzhen Natural Science Foundation- the Stable Support Plan Program (No. 20220810124935001) titled “Research on Key Technologies of Intelligent Detection of Operation Events for Nuclear Power Plants”, the Shenzhen Pengcheng Peacock Plan Talent Project (No. 860-0000020819) titled “Intelligent Analysis Technology of Alarm Flood in Nuclear Power Plants”, Shenzhen Natural Science Foundation- the Stable Support Plan Program (No. 20220811103029001) titled “Research on Risk-informed Dynamic Autonomous Operation Decision-making Technology Based on Real-time Situation Awareness for Nuclear Power Plants”, GuangDong Basic and Applied Basic Research Foundation (No. 2022A1515110545) titled “Research on Performance-based Real-time Risk Monitoring and Dynamic Prediction Technology for Nuclear”, LingChuang Research Project of China National Nuclear Corporation (No. CNNC-LCKY-202263) titled “Research on the reproduction technology of the dynamic interactive operation scenario for the nuclear power plant based on the data mining”, the Shenzhen Pengcheng Peacock Plan Talent Project (No. 827-000885) titled “Visualization technology for dynamic interactive operation scenarios of nuclear power plants based on big data and multi-agent”, Shenzhen Science and Technology Innovation Commission Key Technical Project (JSGG20210713091539014), GuangDong Natural Science Foundation-General Program (No. 2024A1515012727) titled “Research on Dynamic Risk-informed Multi-attribute Objective Collaborative Decision-making Mechanism for Unmanned Small Modular Reactors under Uncertainty Factors”, and Shenzhen Science and Technology Innovation Commission Key Technical Project (JSGG20210713091539014). Professor Ming Yang and Professor Sijuan Chen are the co-corresponding author for this paper.

References

[1] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
[2] Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine[J]. Nature medicine, 2022, 28(1): 31-38.
[3] Zhang C, Han D, Qiao Y, et al. Faster Segment Anything: Towards Lightweight SAM for Mobile Applications[J]. arXiv preprint arXiv:2306.14289, 2023.
[4] He K, Chen X, Xie S, et al. Masked autoencoders are scalable vision learners[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 16000-16009.
[5] Wu J, Fu R, Fang H, et al. Medical SAM adapter: Adapting segment anything model for medical image segmentation[J]. arXiv preprint arXiv:2304.12620, 2023.
[6] Kirillov A, Mintun E, Ravi N, et al. Segment anything[J]. arXiv preprint arXiv:2304.02643, 2023.
[7] Huang Y, Yang X, Liu L, et al. Segment anything model for medical images?[J]. Medical Image Analysis, 2023: 103061.
[8] Cheng J, Ye J, Deng Z, et al. SAM-med2d[J]. arXiv preprint arXiv:2308.16184, 2023.
[9] Wang R, Lei T, Cui R, et al. Medical image segmentation using deep learning: A survey[J]. IET Image Processing, 2022, 16(5): 1243-1267.
[10] Kaul V, Enslin S, Gross S A. History of artificial intelligence in medicine[J]. Gastrointestinal endoscopy, 2020, 92(4): 807-812.
[11] Chen J, Lu Y, Yu Q, et al. Transunet: Transformers make strong encoders for medical image segmentation[J]. arXiv preprint arXiv:2102.04306, 2021.
[12] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.
[13] Wu J, Fu R, Fang H, et al. Medical SAM adapter: Adapting segment anything model for medical image segmentation[J]. arXiv preprint arXiv:2304.12620, 2023.
[14] Zhou H, Gu B, Zou X, et al. A Survey of Large Language Models in Medicine: Progress, Application, and Challenge[J]. arXiv preprint arXiv:2311.05112, 2023.