Marker-Aware Ovarian Tumor Segmentation from Ultrasound Images

Hoang-Son Bui*, Sy-Hoang Tran*, Thuy-Binh Nguyen†, Thanh-Hai Tran*, Hai Vu*, Thi-Lan Le*
* School of Electrical and Electronics Engineering
Hanoi University of Science and Technology, Hanoi, Vietnam
† University of Transport and Communications
Corresponding author e-mail: lan.lethi1@hust.edu.vn

***Abstract*—Ovarian cancer remains one of the leading causes of cancer-related deaths among women, with early detection being pivotal for successful treatment. Accurate segmentation of ovarian tumor regions in ultrasound images is essential to assist clinicians in the effective diagnosis and treatment of ovarian cancer. However, many previous studies have struggled to achieve high segmentation accuracy due to the complex nature of ultrasound images. These images often contain inherent noise, and symbols marked by sonographers, which pose significant challenges for segmentation tasks. In this study, before feeding the images into the segmentation model, we employ an in-painting method to remove the symbols and markers from the ultrasound image. Then a segmentation model based on UNet3+ with ResNet50 as its encoder is introduced. We evaluated the performance of our model on the OTU2D dataset and demonstrated an improvement over the existing models with the Dice, IoU, Recall and Precision scores of 86.44%, 77.05%, 86.31%, and 89.18%, respectively. Moreover, when testing on a subset of clean images, the pre-processing technique based on in-painting helps to increase the Dice and Precision metrics from 90.63% and 86.23% to 96.64% and 96.08%. This study presents a novel approach that enhances image segmentation capabilities by introducing an in-painting pre-processing method and a UNet3+ model with ResNet50 as its encoder. Our method significantly improves segmentation accuracy and offers clinicians more reliable data for better-informed clinical decisions.**

***Index Terms*—Ovarian Tumor Segmentation; In-painting; Marker Removal; OTU2D Dataset; ResNet50-UNet3+;**

I. Introduction

Ovarian cancer progresses silently and is often diagnosed at late stages, making it a leading cause of cancer death in women. Previous studies show that if cancer is diagnosed early, the five-year survival rate exceeds 90% [1]. Early detection and accurate diagnosis of the tumors are crucial for timely treatment, which can significantly improve patient outcomes and maintain women's health and fertility. Identifying the tumor area can assist doctors in predicting and deciding appropriate treatment methods, thereby improving the overall quality of healthcare services.

Determining the tumor region of interest from ultrasound images is often formulated as a segmentation task. Segmentation tasks in medical imaging often utilize CNNs with encoder-decoder architectures to extract relevant features from images [2], [3]. Typically, the input of these models is an ovarian ultrasound image, and the output is a binary image in which tumor areas are separated from non-tumor areas. However, the complex nature of ultrasound images, with their inherent noise and the symbols marked by sonographers, poses significant challenges for ovarian tumor segmentation.

In our study, to overcome the challenges posed by noise and symbols in ultrasound images, we utilized the LaMa in-painting technique [4] as a pre-processing step. This approach effectively eliminates artifacts such as sonographer markers, measurement rulers, and patient details, enhancing the image quality and facilitating more accurate segmentation when applied to deep learning models. Then a segmentation model that couples the Resnet50 encoder with the architecture of UNet3+ is introduced. The results of our model on the OTU2D dataset [5] demonstrate an improvement over the existing models with the Dice, IoU, Recall, and Precision scores of 86.44%, 77.05%, 86.31%, and 89.18%. Moreover, when testing on a subset of 50 clean images, the pre-processing technique based on in-painting helps to increase the Recall and Precision metrics from 89.32% and 86.23% to 97.78% and 96.08%.

In summary, the contribution of the paper is two-fold. First, we propose to apply an in-painting technique in the pre-processing step to remove the unwanted markers and symbols in ultrasound images. Second, a ovarian tumor method that incorporates the ResNet50 encoder and the architecture of Unet3+ with full-scale skip connections is proposed.

The structure of this paper is as follows: Section II provides a brief review of existing in-painting techniques and ovarian tumor segmentation methods. Section III describes the main components of our proposed framework. The evaluation results are discussed in Section IV. Finally, Section V offers discussions and conclusions.

II. Related Work

Segmentation models have seen significant advancements recently, with a notable increase in research outcomes. Ronneberger et al. introduced U-Net [6], which is a widely adopted architecture for biomedical image segmentation. Its architecture includes a contracting path for capturing contextual information and an expanding path for precise localization. Skip connections between corresponding layers to preserve fine details. U-Net excels at handling small datasets and accurately segmenting medical images by capturing intricate details and spatial context simultaneously.

Building upon [6], U-Net++, an extension of UNet, was developed by Zhou et al. in [7]. UNet++ aims to improve segmentation accuracy. It integrates nested and densely connected pathways to capture hierarchical features across multiple scales. U-Net++ demonstrates enhanced performance by leveraging nested skip pathways to refine segmentation boundaries and details. Additionally, U-Net3+ [8], incorporates attention mechanisms and dense connectivity blocks to enhance feature reuse and integration across scales. This architecture achieves superior segmentation results by focusing on effective feature aggregation and refinement at different hierarchy levels.

To enhance segmentation capabilities in ultrasound images, in [9], a method named CR-Unet was proposed. CR-Unet integrates a spatial recurrent neural network (RNN) with a U-Net architecture, forming a composite network. CR-Unet addresses several typical challenges of ultrasound images, such as poor image quality, low contrast, and complex anatomical shapes.

Recent studies have tried to enhance UNet by [6] not only optimizing skip connections but also improving the encoder branches. For instance, the study in [2] replaced the encoder branch by MobileNetV2 to enhance feature extraction from ultrasound images. This approach led to a significant increase in the Dice coefficient to 79.00% on the OTU2D dataset [5]. It shows promising results in semantic segmentation tasks, demonstrating the effectiveness of integrating advanced encoder architectures within the UNet framework. Additionally, it highlights the potential of using more powerful architectures to further enhance model performance.

Several studies focus on addressing noise in ultrasound images, such as patient details and markers. In [1], Coburn et al. employed a CNN-CAE approach to diagnose ovarian tumors. They initially faced challenges with misaligned activation areas in their CNN model due to marks present in the images. By incorporating a denoising convolutional autoencoder (CAE) to remove these marks before training the CNN, they achieved improved classification results. This adjustment enabled more accurate alignment of activation areas with relevant regions in the images, thereby facilitating more reliable tumor diagnosis.

In [10], Chen et al. focused on enhancing classification and segmentation accuracy by addressing symbols and marks in medical images using image in-painting techniques. They introduced a framework called mask-guided generative adversarial network (MGGAN) designed to remove these symbols. MGGAN utilizes attention mechanisms to enhance the realism of lesion boundaries, thereby improving segmentation accuracy. Their approach significantly boosted segmentation performance, increasing accuracy from 71.51% to 76.06% for the Unet model and from 61.13% to 66.65% for the PSPnet model when applied to clean images. These studies underscore the importance of robust pre-processing techniques in medical image analysis, demonstrating how advancements in image cleaning and enhancement can lead to more accurate and reliable AI-based diagnostic tools.

III. Proposed Framework for Ovarian Tumor Segmentation

A. Overall framework

The main target of this study is to propose an effective framework for the ovarian tumor segmentation task, illustrated in Fig. 1. This framework consists of two crucial components that are pre-processing and image segmentation steps. In fact, ultrasound images usually contain inherent noise and symbols marked by sonographers, leading to significant challenges for segmentation tasks. Consequently, the purpose of the preprocessing step is to eliminate noise and provide higher-quality images for the subsequent segmentation step. For the image segmentation task, we leverage the abilities of both the ResNet50 [11] and U-Net3+ [6] models by replacing the encoder of U-Net3+ with ResNet50 for the feature extraction step. The ResNet50 model enhances the segmentation capability of U-Net3+, significantly improving both Precision and Recall in the ovarian tumor segmentation problem. The details of these two crucial steps are provided in the following sections.

B. Symbol removing by in-painting technique

With the purpose of eliminating noise and providing higher images for the segmentation step, in this study, we propose to use the Large-Mask in-painting technique (LaMa) [4] in the pre-processing step. The LaMa technique is performed to remove and fill up large missing areas within a given image; from this, the input images become clearer and sharper. However, this process is time-consuming, particularly during the labeling phase, where annotating noise markers in ovarian ultrasound images. The Fast Fourier Convolution (FFC) block is a key component in this model, designed to split channels into local and global information branches in images. While real FFT (Fast Fourier Transform) is utilized in the global branch, enabling the model to take the entire image into account when generating convolutions, it is applied to real-valued signals, ensuring that the output remains real-valued. A local branch utilizes conventional convolutions to process information within specific regions of the image. By combining the outputs from their local and global branches, FFCs effectively leverage both local details and global context for accurate image in-painting. This approach provides an image-wide receptive field, enabling the generator to integrate global context from the initial layers, which is crucial for ensuring the filled areas remain consistent with the overall structure and appearance of the original image. The architecture of the LaMa model is described in Fig. 2.

C. Ovarian Tumor Segmentation

The main contribution of the proposed framework is demonstrated in the image segmentation component, where the ResNet50 [11] model is incorporated with U-Net3+ to enhance its segmentation capability. Fig.3 illustrates the modification in the architecture of U-Net3+, with ResNet50 serving as the encoder of the U-Net3+ model. By utilizing ResNet50

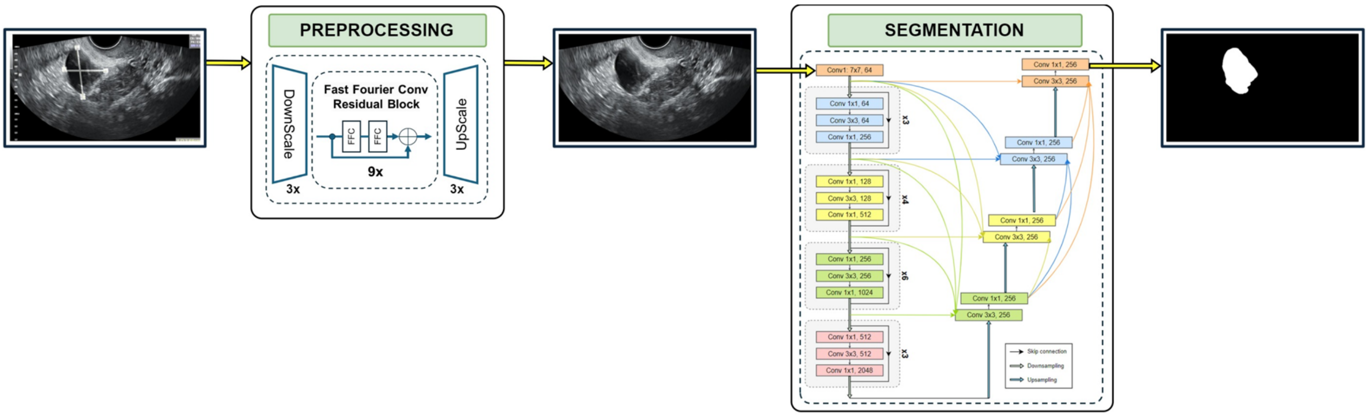

Fig. 1. The proposed framework for ovarian tumor segmentation.

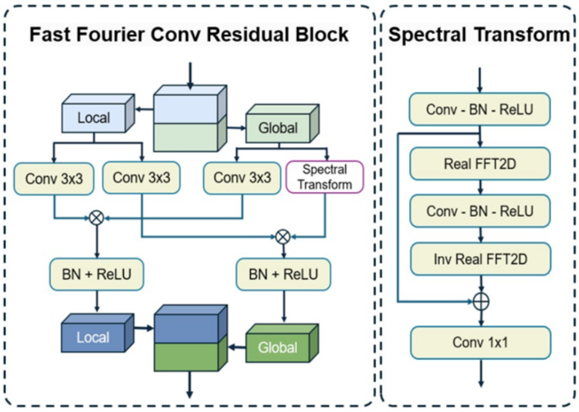

Fig. 2. Large Mask In-painting architecture.

as a pre-trained encoder, we leverage its ability to efficiently extract meaningful features from input images, significantly reducing both training time and computational demands compared to building a model from scratch. This approach is particularly advantageous because training from scratch can be resource-intensive and expensive. By leveraging the pre-trained ResNet50 encoder, we benefit from its proven performance and robustness, thereby improving the overall effectiveness and efficiency of the ovarian tumor segmentation task.

The decoder remains consistent with the original U-Net architecture. Integrating the U-Net architecture with ResNet50 has proven to be highly effective for segmentation tasks compared to the standard U-Net. The combination of U-Net and ResNet50 leverages the power of both models: the detailed spatial information captured by U-Net and the deep feature extraction capability of ResNet50. This synergy enhances the accuracy and efficiency of the segmentation process, making it particularly suitable for complex medical imaging tasks such as ovarian tumor detection and delineation. By maintaining the foundational elements of U-Net while incorporating the advanced features of ResNet50, this approach not only improves performance but also ensures that the segmentation process is more reliable and precise.

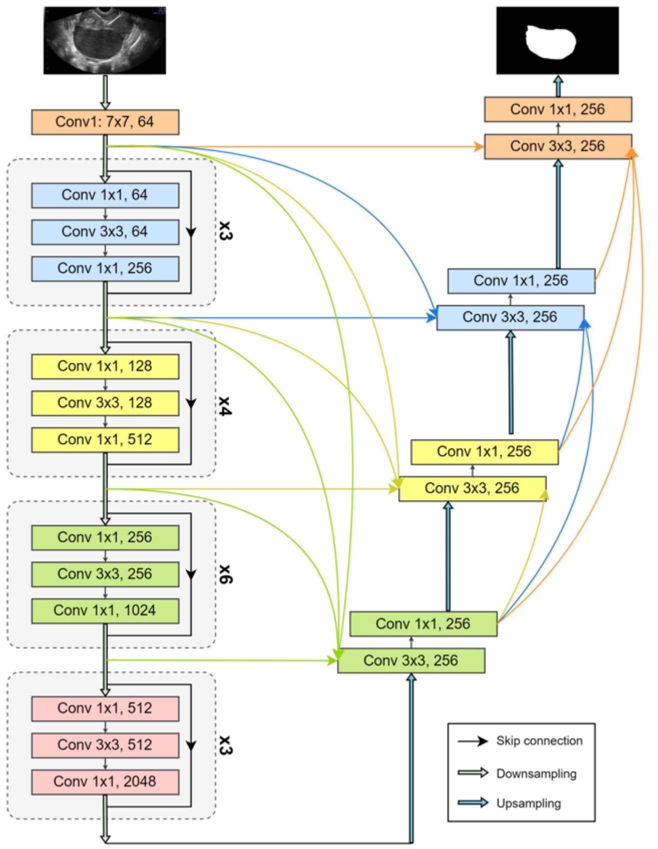

Fig. 3. The segmentation method incorporating ResNet50 and U-Net3+ architecture.

D. Loss function

For this study, in our experiment, we employ Binary Cross Entropy (BCE) [12], which is particularly suitable for binary segmentation tasks. The BCE is defined as follows:

$$\mathrm{BCE}(y, \hat{y}) = -\left(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})\right) \quad (1)$$

where y represents the ground truth area of the ovary in ultrasound images, and $\hat{y}$ denotes the region of the ovary predicted by the model.

IV. Experiments and Results

A. Dataset

We utilized a benchmark dataset, OTU2D, which comprises 1469 ovarian ultrasound images collected from 294 patients at Beijing Shijitan Hospital and Capital Medical University, China [5]. However, in this study, we focus on binary segmentation to identify the presence or absence of tumors. To ensure a fair evaluation, we followed the split ratio protocol 8:1:1 images for training, testing, and validation, respectively. Figure 4 illustrates several sample images from the OTU2D dataset, showcasing the diverse types of tumors.

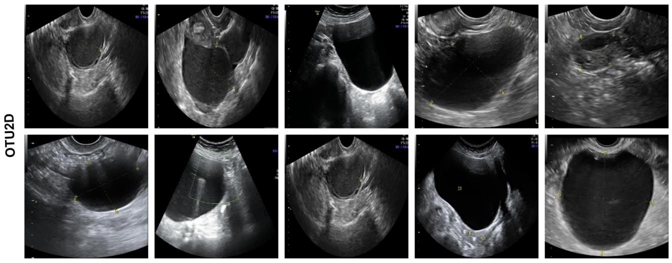

Fig. 4. Illustration of several ovarian tumor ultrasound images.

B. Experimental results

1) Qualitative evaluation of preprocessing step: Figure.5 illustrates some results of the preprocessing step. We can observe that markers, symbols, and other annotations have been removed. Although in some parts the contrast level is decreased, we can still see the detailed information of different parts of the ovarian area.

2) Ablation Study: To evaluate the role of each component in the proposed method, in our study, we have conducted three experiments. The first experiment aims to evaluate the performance of UNet3+, while the second experiment tried to assess the robustness of the encoder part. The final experiment is to compare the BCE loss function with other loss functions.

Table I shows the results of the first experiment. Among the three architectures that are UNet, UNet++, and UNet3+, UNet3+ demonstrates the best performance when using skip connections with Dice (84.6%), IoU (74.32%), and Recall (87.42%). Although UNet3+ outperforms UNet++ and UNet in terms of dice, IoU, and recall, UNet demonstrates higher precision (86.43%) in predicting the pixels of the regions belonging to ovarian tumors compared to UNet3+ (0.54%), with precision lower by 0.54% compared with UNet. The findings highlight a notable advantage when utilizing skip-connections to interconnect layers within UNet3+, leading to significantly superior performance compared to both UNet and UNet++ architectures.

In the second experiment, we compare the performance of three different encoders, which are VGG16, MobileNetV2, and ResNet50. Results are shown in Tab. II. ResNet50 outperforms VGG16 and MobileNetV2 in Dice (86.44%), IoU (77.05%), and Recall (89.18%), highlighting its effectiveness

TABLE I

Comparison of UNet, UNet++ and UNet3+ backbones for ovarian tumor segmentation. Best results are in bold (%).

Model	Dice	IoU	Recall	Precision
UNet [6]	81.02	69.44	83.76	**86.43**
UNet++ [7]	73.68	59.34	79.07	82.55
UNet3+ [8]	**84.62**	**74.32**	**87.42**	85.89

for high-quality segmentation. VGG16 has the highest Recall (88.11%) but is missing many important information. Fig. 6 shows results from different backbones, demonstrating that the proposed method closely matches the ground truth.

TABLE II

Comparison of encoders when incorporating with UNet3+ architecture. Best results are in bold (%).

Model	Dice	IoU	Recall	Precision
MobileNetV2 + UNet3+	77.74	64.52	78.32	82.08
VGG16-Net + UNet3+	85.36	75.74	**88.11**	87.15
ResNet-50 + UNet3+	**86.44**	**77.05**	86.31	**89.18**

Loss function is also important in the segmentation model. Our third experiment aims at assessing the performance of loss functions. In the literature, there are several loss functions proposed for ovarian tumor segmentation. Table III highlights a comparison of the performance of various loss functions. The proposed method utilizing BCE loss stands out with Dice (86.44%) and IoU (77.05%), reflecting the effectiveness of the proposed model with BCE. In comparison, Dice loss performs well but is slightly lower than that achieved with BCE loss. IoU Loss shows the highest recall (89.29%) but a low precision (85.51%), indicating that it misses many important regions within the ovarian tumor region. Focal loss excels in precision (90.90%) but low recall (83.24%) and IoU (70.17%) show that it misses many other important regions that we need to segment. The Hybrid loss [2] and Joint loss [3] methods provide a balanced performance but do not surpass BCE loss in any metric. These results in Fig. 7 suggest that BCE loss is the most effective for our proposed method, achieving a good balance between precision and recall while maintaining high accuracy.

TABLE III

Comparison of loss functions. Best results are in bold (%).

Loss	Dice	IoU	Recall	Precision
BCE	**86.44**	**77.05**	86.31	89.18
Dice [13]	85.70	76.11	84.83	88.97
IoU [14]	86.14	76.61	**89.29**	85.51
Focal [15]	81.71	70.17	83.24	**90.90**
Hybrid [2]	85.67	75.89	83.97	89.89
Joint [3]	85.03	75.19	84.85	88.90

3) Comparison with the state-of-the-art methods: Table IV compares the segmentation results of the proposed method with the state-of-the-art methods on the OTU2D dataset. We can observe that the proposed method using BCE Loss achieves better performance than MU-Net [2].

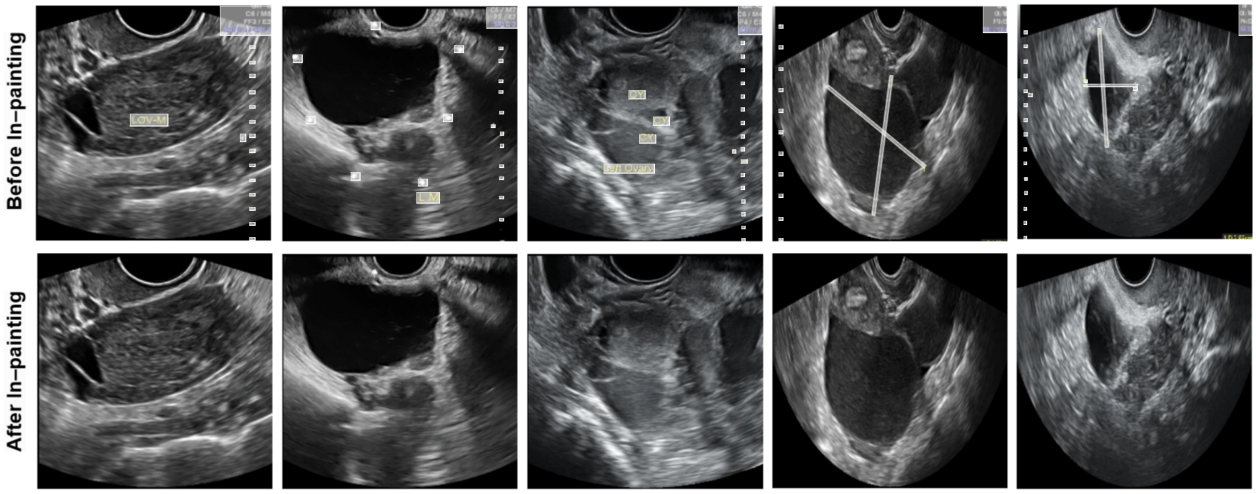

Fig. 5. Example of ultrasound images before an after preprocessing step in the OTU2D dataset. The first row shows the ultrasound images with markers highlighted by bounding boxes, while the second row shows the preprocessed images where markers and symbols are removed.

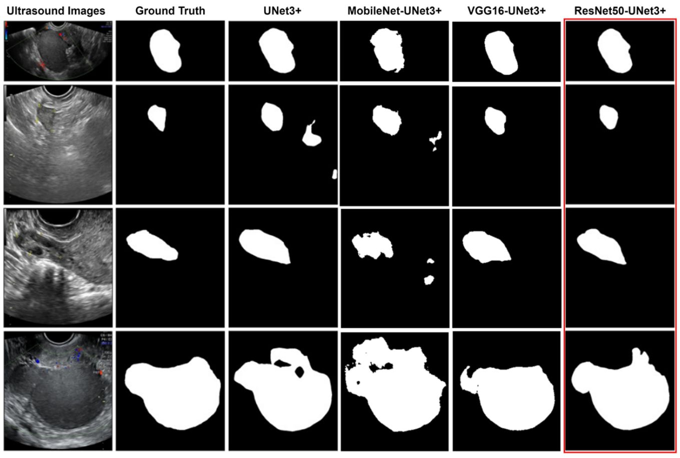

Fig. 6. Example of segmentation results from difference models. Results of the proposed result are highlighted in a red rectangle.

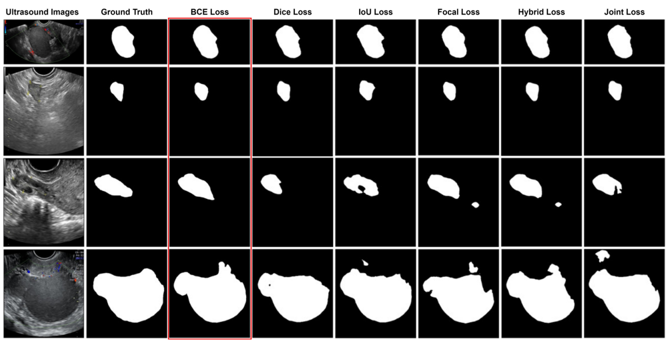

Fig. 7. Example of segmentation results from difference metrics loss. Results of the proposed method are highlighted in a red rectangle.

TABLE IV
COMPARISON WITH THE STATE-OF-THE-ART METHODS ON OTU2D DATASET (%)

Model	Loss	Dice	IoU	Recall	Precision
MU-Net [2]	BCE	56.70	39.80	77.20	82.70
MU-Net [2]	Hybrid	79.00	65.00	80.00	82.00
Proposed Method	BCE	**86.44**	**77.05**	**86.31**	**89.18**

To evaluate the effect of marker removal, 50 clean images from the OUT2D dataset were used. Two scenarios are defined using 1419 marker images for training. In the first scenario (without preprocessing), we employ the original images, while in the second scenario (with preprocessing), output images of LaMa models are utilized. Experimental results are reported in Tab. V. The results confirm the impact of symbol and marker removal in the preprocessing step, with the values of Dice, IoU, Recall, and Precision metrics being 96.6%, 93.57%, 97.78%, and 96.08%, respectively. Fig.8 illustrates some segmentation results.

TABLE V
COMPARISON RESULTS OF USING THE PROPOSED METHOD OF TESTING ON THE CLEAN IMAGES SUBSET IN THE OTU2D DATASET (%)

Scenario	Dice	IoU	Recall	Precision
Without preprocessing	90.63	83.02	89.32	86.23
With preprocessing	**96.64**	**93.57**	**97.78**	**96.08**

V. CONCLUSIONS AND FUTURE WORKS

In this study, we have proposed an effective framework for ovarian tumor segmentation by incorporating a pre-processing step and a segmentation network based on ResNet50 and U-Net3+. Several experiments were conducted on a benchmark dataset, demonstrating significant improvements, with the Dice Coefficient getting 86.64% on the OTU2D dataset and increasing from 90.63% to 96.64% when using the LaMa in-painting technique. Compared to previous studies, which often focus

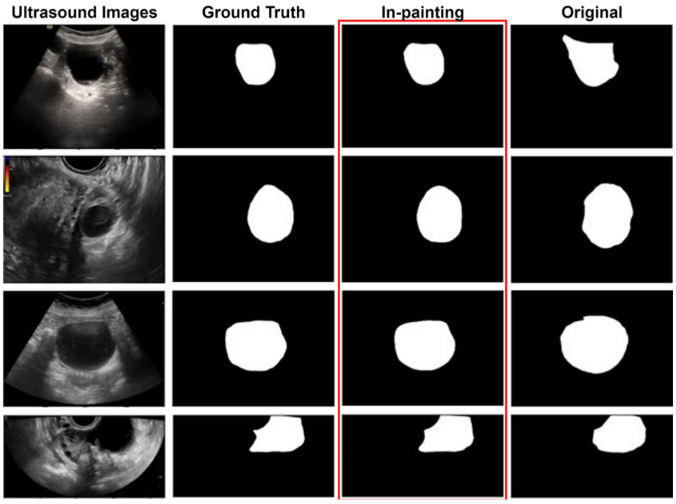

Fig. 8. Example of segmentation results with and without preprocessing step. Results with preprocessing step are highlighted in a red rectangle.

on enhancing segmentation performance without addressing the impact of annotations and noise, our work offers a more comprehensive approach by combining pre-processing with segmentation techniques. Our study contributes to the field by addressing this challenge, thus improving the performance of the segmentation tasks. However, despite the notable performance, our model still has limitations in the boundary. The future work will focus on refining these boundaries and leveraging the segmented regions to compute clinically relevant parameters, such as orthogonal diameters or boundary features.

Acknowledgement

This research is funded by the Ministry of Science and Technology of VietNam (MOST) under grant number KC-4.0-45/19-25 **"Research and development of a computer-aided support system for ovarian cancer diagnosis using ultrasound images"**.

References

[1] S. B. Coburn, F. Bray, M. E. Sherman, and B. Trabert, "International patterns and trends in ovarian cancer incidence, overall and by histologic subtype," *International Journal of Cancer*, vol. 140, no. 11, pp. 2451–2460, 2017. DOI: 10.1002/ijc.30676.

[2] T.-H. Nguyen, T.-L. Pham, Q.-V. Tran, *et al.*, "Systematic evaluation of loss functions for ovarian tumors segmentation from ultrasound images," in *2023 1st International Conference on Health Science and Technology (ICHST)*, 2023, pp. 1–6. DOI: 10.1109/ICHST59286.2023.10565336.

[3] H.-P. Luong, H.-S. Bui, N.-K. Nguyen, *et al.*, "Sovasegnet: Scale invariant ovarian tumors segmentation from ultrasound images," in *2024 IEEE International Conference on Image Processing (ICIP)*, 2024, pp. 2081–2087. DOI: 10.1109/ICIP51287.2024.10647995.

[4] R. Suvorov, E. Logacheva, A. Mashikhin, *et al.*, "Resolution-robust large mask inpainting with fourier convolutions," *arXiv preprint arXiv:2109.07161*, 2021.

[5] Q. Zhao, S. Lyu, W. Bai, *et al.*, "A multi-modality ovarian tumor ultrasound image dataset for unsupervised cross-domain semantic segmentation," *arXiv preprint arXiv:2207.06799*, 2022.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention 2015*, Cham: Springer International Publishing, 2015, pp. 234–241, ISBN: 978-3-319-24574-4.

[7] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," pp. 3–11, 2018.

[8] H. Huang, L. Lin, R. Tong, *et al.*, "Unet 3+: A full-scale connected unet for medical image segmentation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 1055–1059. DOI: 10.1109/ICASSP40776.2020.9053405.

[9] H. Li, J. Fang, S. Liu, *et al.*, "Cr-unet: A composite network for ovary and follicle segmentation in ultrasound images," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 974–983, 2020. DOI: 10.1109/JBHI.2019.2946092.

[10] L. Chen, C. Qiao, M. Wu, *et al.*, "Improving the segmentation accuracy of ovarian-tumor ultrasound images using image inpainting," *Bioengineering*, vol. 10, no. 2, p. 184, 2023, ISSN: 2306-5354. DOI: 10.3390/bioengineering10020184.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[12] S. Jadon, "A survey of loss functions for semantic segmentation," in *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2020, pp. 1–7. DOI: 10.1109/CIBCB48159.2020.9277638.

[13] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Cham: Springer International Publishing, 2017, pp. 240–248, ISBN: 978-3-319-67558-9.

[14] J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, "Unitbox: An advanced object detection network," in *Proceedings of the 24th ACM international conference on Multimedia*, ACM, 2016, pp. 516–520.

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *arXiv preprint arXiv:1708.02002*, 2017. DOI: 10.48550/arXiv.1708.02002. arXiv: 1708.02002 [cs.CV].