scientific reports

Check for updates

OPEN

A weakly-supervised follicle segmentation method in ultrasound images

Guanyu Liu[1], Weihong Huang[1,2,3,9], Yanping Li[4,5,9], Qiong Zhang[4,5,9], Jing Fu[4,5,9], Hongying Tang[4,5,9], Jia Huang[6,9], Zhongteng Zhang[7,9], Lei Zhang[8,9], Yu Wang[4,5]✉ & Jianzhong Hu[1,2]✉

Accurate follicle segmentation in ultrasound images is crucial for monitoring follicle development, a key factor in fertility treatments. However, obtaining pixel-level annotations for fully supervised instance segmentation is often impractical due to time and workload constraints. This paper presents a weakly supervised instance segmentation method that leverages bounding boxes as approximate annotations, aiming to assist clinicians with automated tools for follicle development monitoring. We propose the Weakly Supervised Follicle Segmentation (WSFS) method, a novel one-stage weakly supervised segmentation technique model designed to enhance the ultrasound images of follicles, which incorporates a Convolutional Neural Network (CNN) backbone augmented with a Feature Pyramid Network (FPN) module for multi-scale feature representation, critical for capturing the diverse sizes and shapes of follicles. By leveraging Multiple Instance Learning (MIL), we formulated the learning process in a weakly supervised manner and developed an end-to-end trainable model that efficiently addresses the issue of annotation scarcity. Furthermore, the WSFS can be used as a prompt proposal to enhance the performance of the Segmentation Anything Model (SAM), a well-known pre-trained segmentation model utilizing few-shot learning strategies. In addition, this study introduces the Follicle Ultrasound Image Dataset (FUID), addressing the scarcity in reproductive health data and aiding future research in computer-aided diagnosis. The experimental results on both the public dataset USOVA3D and private dataset FUID showed that our method performs competitively with fully supervised methods. Our approach achieves performance with mAP of 0.957, IOU of 0.714 and Dice Score of 0.83, competitive to fully supervised methods that rely on pixel-level labeled masks, despite operating with less detailed annotations.

Keywords Weakly Supervised Learning, Ultrasound Image Segmentation, Assisted Reproductive Technology

Assisted reproduction technology (ART)[1] is a beacon of hope for millions of people and couples around the world grappling with the daunting challenge of infertility. Through a myriad of innovative techniques, ART has revolutionized the landscape of reproductive medicine, offering pathways to parenthood that were once considered unattainable.

Among these techniques, follicle growth monitoring is crucial for optimizing treatment protocols and increasing the chances of successful outcomes. Clinicians have relied on ultrasound imaging to assess ovarian follicles and determine their readiness for interventions such as controlled ovarian stimulation (COS)[2]. In this context, the measurement of the maximum diameter of the largest follicle emerges as a critical metric, serving as a key indicator of follicular maturity and guiding treatment decisions. However, reliance on manual assessment

[1]Big Data Institute, Central South University, Changsha 410083, China. [2]Mobile Health Ministry of Education - China Mobile Joint Laboratory, Xiangya Hospital, Central South University, Changsha 410000, China. [3]Xiangjiang Laboratory, Changsha 410205, China. [4]Department of Reproductive Medicine, Xiangya Hospital, Central South University, Changsha, Hunan 410000, China. [5]Clinical Research Center for Women's Reproductive Health in Hunan Province, Changsha, Hunan 410000, China. [6]School of Life Science, Central South University, Changsha 410083, China. [7]School of Computer Sciences and Engineering, Central South University, Changsha 410083, China. [8]Laboratory of Vision Engineering (LoVE), School of computer science, University of Lincoln, Lincoln LN6 7TS, UK. [9]Weihong Huang, Yanping Li, Qiong Zhang, Jing Fu, Hongying Tang, Jia Huang, Zhongteng Zhang and Lei Zhang contributed equally to this work. ✉email: 541214404@qq.com; jianzhonghu@csu.edu.cn

methods introduces inherent limitations, as their outcomes can vary depending on the experience and skill level of the individual physician.

To address the challenges of manual assessment and mitigate potential human errors in follicle evaluation, the integration of computer-aided tools presents a promising solution. In follicle assessment, image segmentation plays a crucial role in accurately delineating the boundaries of the follicles. By leveraging computational algorithms and machine learning, these tools have the potential to augment and enhance clinician capabilities, facilitating more accurate and objective assessments of follicle development.

Despite significant advancements in medical image segmentation methods[3–5], there remains a lack of applications in the field of reproductive medicine, due to 1) limited availability of public follicle ultrasound image datasets, 2) the difficulty of obtaining pixel-level annotations, which requires significant time and expertise, and 3) ultrasound images inherently possess a high-noise characteristic[6,7], which poses a major challenge for segmentation. Additionally, the individual follicles vary in size, and there is a possibility that they may obscure one another. These factors further exacerbate the difficulty of the segmentation task.

Previous research in weakly supervised instance segmentation has explored various approaches to the ultrasound image segmentation work for a real-time medical system[8]. However, fully supervised instance segmentation models, while highly effective, require extensive labeled data to achieve their potential, making them challenging to deploy due to the scarcity of annotated images. In contrast, weakly supervised instance segmentation approaches, which operate with limited or no object-specific annotations, offer a practical and more feasible alternative. These methods have made significant strides, yet they still face performance gaps compared to their fully supervised counterparts.

In this study, we introduce a novel weakly supervised learning approach for segmenting follicle ultrasound images, utilizing a Convolutional Neural Network (CNN)[9–12] backbone. Our method uses bounding boxes as annotations to train the model, tackling the segmentation task in a clinical context where obtaining precise pixel-level ground truth (GT) is often challenging. We proposed an end-to-end trainable model that is learned in a weakly supervised manner incorporating the principles of multiple instance learning (MIL)[13]. MIL is a deep learning paradigm where training instances are grouped into bags, with the assumption that at least one instance in each positive bag is relevant to the task, allowing for the handling of uncertain labels and unsupervised instance-level information.

Furthermore, pre-trained models such as the Segment Anything Model(SAM)[14], necessitate user prompts and fine-tuning to adapt to specific domains like medical imaging. The requirement for prompt-based interaction and the computational resources needed for fine-tuning can limit the applicability of SAM in resource-constrained environments or scenarios where direct application to medical images without fine-tuning is desired. Our work aims to fill this gap by adapting weakly supervised methods to the unique characteristics of follicle ultrasound images. We also aim to improve the SAM-like architecture by introducing optimized prompts with the weakly supervised pipeline, rather than conducting adjustments in the post-pre-training tuning stage. More specifically, WSFS can also be plug-and-played to SAM-Med2D[15], an adaptation of the Segment Anything Model (SAM)[14] for medical images, leveraging similar principles but tailored to the unique characteristics and challenges of medical imaging data.

The main contributions of this work are summarized as follows. First, we proposed a weakly supervised one-stage instance segmentation method using coarse annotations on follicle ultrasound images, which can be further integrated with SAM-Med2D to enhance final segmentation results. To the best of our knowledge, this is the first weakly supervised follicle segmentation method in ultrasound images specifically designed to address the issue of limited GT data in this field. Second, we created a Follicle Ultrasound Image Dataset (FUID) to address the current data shortage in assisted reproductive research.Our method demonstrates substantially competitive performance compared to popular instance segmentation methods and the state-of-the-art fully supervised instance segmentation approach.

Related works

Deep learning methods on ovarian ultrasound image

Deep learning[16] has been increasingly applied to ovarian ultrasound imaging to improve the detection and classification of ovary and follicles. Early works in this area[9,17–20] focused on manual feature extraction and traditional machine learning techniques to classify tumors in ultrasound images. With the advent of deep learning, CNNs have been employed to automatically learn high-level features from ultrasound images, leading to significant improvements in classification accuracy. For instance, 3D CNNs[11,21]have been utilized to process volumetric ultrasound data for more comprehensive tumor characterization. Moreover, researchers have delved into attention mechanisms and hybrid architectures that integrate features from diverse modalities to boost the performance of ovarian tumor segmentation and classification. However, despite these progressions, the implementation of deep learning techniques in ovarian ultrasound imaging still encounters obstacles. The appearance variability of follicles or lesions (tumors) poses a significant challenge, and due to the data - intensive nature of deep learning, the performance is highly dependent on the availability of large, annotated datasets.

Fully supervised image segmentation

Fully supervised image segmentation aims to delineate each instance of a foreground object class in an image, using a large amount of annotated data for training. This approach has been widely adopted in medical image analysis, including the segmentation of tumors in ultrasound images. Recent studies have leveraged powerful deep learning architectures, such as U-Net[10] and its variants[11,17–19,21,22], to achieve high-resolution segmentation of ovarian structures in ultrasound images. These methods typically require pixel-level annotations for each instance, which can be labor-intensive to obtain. You Only Look Once (YOLO)[23–32] and DeepLab[33–38] are two prominent model series, both of which have seen multiple iterations with incremental improvements.

For segmentation, in YOLOV8[31], YOLO-seg was introduced as an extension of the YOLO object detection framework that incorporates segmentation capabilities. For DeepLab, this series of models has been pivotal in semantic image segmentation, utilizing deep convolutional neural networks coupled with Conditional Random Fields (CRFs)[33] to refine segmentation results. Despite the success of fully supervised methods, YOLO-seg and DeepLab models face limitations in generalizing to small or overlapping objects and often sacrifice some accuracy for the sake of speed. As typical supervised methods, they usually requires pixel-wise annotations for training. Mask R-CNN[12], which extends Faster R-CNN[39], incorporates an additional branch dedicated to predicting segmentation masks for each Region of Interest (RoI). This enhancement enables instance segmentation, allowing for the simultaneous generation of both bounding boxes and segmentation masks for every detected object. Path Aggregation Network (PANet)[40] is an enhancement to the Faster R-CNN[39] model that improves feature extraction by aggregating features from multiple paths. It helps in capturing both global and local context, leading to better detection performance. SOLOv1[41] and SOLOv2[42] is a novel instance segmentation framework called "Segmenting Objects by Channels", where v1 segments objects by predicting object masks in separate channels and classifying them, providing a new perspective on instance segmentation that is efficient and accurate. SOLOv2 refines the instance segmentation approach, enhancing the performance and robustness of the algorithm with better mask representation and classification. You Only Look At Transformed Coordinates (YOLACT)[43] is an instance segmentation method that builds on YOLOv3[25]. It segments objects by predicting masks in a transformed coordinate space, allowing for real-time instance segmentation. YOLACT++[44], an advancement over YOLACT, improves on its predecessor by refining the mask prediction process and enhancing the overall performance, offering better accuracy and efficiency in instance segmentation tasks.

Weakly supervised instance segmentation

The field of weakly supervised instance segmentation has seen significant advancements with the introduction of various models that leverage limited annotation data to achieve competitive performance. A key development in this area is the transition from models like Bounding Box Tightness Prior (BBTP)[13] to more advanced approaches such as Box2Mask[45]. BBTP introduced a multiple instance learning (MIL) framework that utilized the tightness of bounding boxes to generate positive and negative bags for training. This method was among the early ones to demonstrate that weak supervision, presented in the form of bounding boxes, could be effectively applied to instance segmentation tasks. However, BBTP was not end-to-end trainable and relied on post-processing techniques such as GrabCut[46] and Multiscale Combinatorial Grouping (MCG)[47] proposals to generate pseudo GT, which limited its performance. Building upon the foundation laid by BBTP, BoxInst[48] employed a color-pairwise affinity with box constraints in a region-based framework, improving the end-to-end training capability and reducing the dependency on pseudo GT generation. BoxInst demonstrated better performance by directly learning from the box annotations without the need for intermediate proposals. DiscoBox[49] made significant progress in the field. It integrated intra-image and cross-image pairwise potentials and constructed a self-ensembled teacher network. This teacher network generated pseudo-labels, which were then utilized to direct the learning process of the student network. This method showed improvements by considering the relationships between pixels both within and across images. And Box2Mask[45], presents a novel single-shot instance segmentation approach that integrates the classical level-set evolution model into deep neural network learning. This approach achieves accurate mask prediction with only bounding box supervision by evolving level-set curves implicitly using both the input image and its deep features.

Zero-shot/Few-shot instance segmentation

The pursuit of zero-shot instance segmentation has driven the development of methods that can discern and segment objects without explicit training on the target classes. This is especially pertinent in medical imaging, where annotated data can be limited or non-existent for specific conditions or clinical environments. Zero-shot learning approaches have emerged as a promising solution to this problem, relying on the transference of knowledge from a source domain rich with annotations to a target domain that lacks annotations. Common strategies in zero-shot learning include domain adaptation, where models are trained to adapt from a different but related domain, feature alignment, which seeks to align the feature spaces of the source and target domains, and generative models that generate plausible instances of the target classes. The Segment Anything Model (SAM)[14] provides an interactive image segmentation framework that allows users to segment images by providing minimal input, such as points, bounding boxes or masks. SAM is designed to be versatile and can handle a wide range of segmentation tasks. The model is pre-trained on a large dataset and can generalize well to new images and scenarios. The SAM-Med2D[15], an adaptation of SAM for medical images, leverages similar principles but is specifically designed to address the unique characteristics and challenges of medical imaging data. It incorporates domain-specific knowledge and constraints to improve segmentation accuracy and robustness. The adaptability of SAM-Med2D makes it a valuable tool for zero-shot image segmentation in medical imaging, as it can potentially generalize across different types of medical images and pathologies.

Proposed methods

In this section, we introduce the proposed method. First, we provide an overview of the approach, followed by a detailed description of the weakly supervised instance segmentation method specifically designed for follicle ultrasound images. Finally, we discuss the post-processing or segment refinement techniques and implementation details.

Overview

We propose a novel one-stage weakly supervised segmentation method called Weakly Supervised Follicle Segmentation (WSFS), designed for the segmentation of ultrasound images of follicles. This method addresses

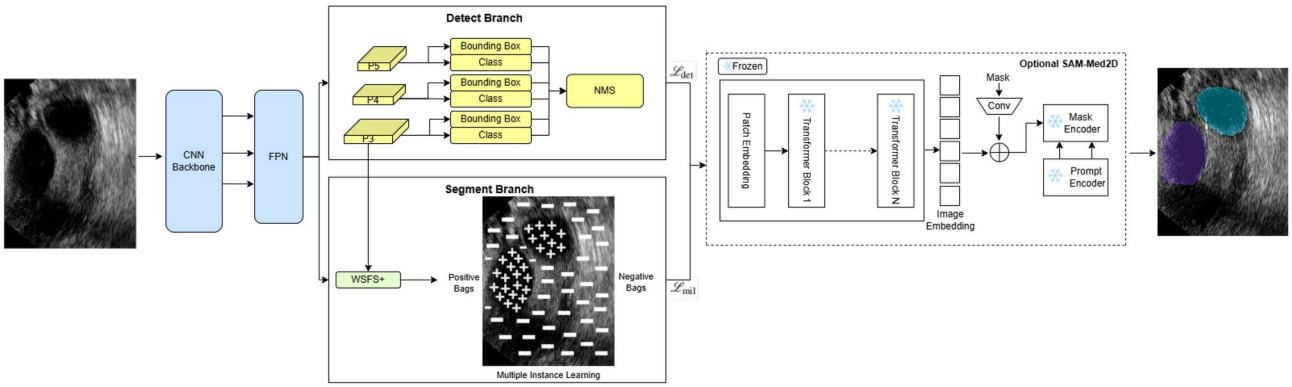

Fig. 1. An overview of the proposed method, Weakly Supervised Follicle Segmentation (WSFS). WSFS consists of a convolutional neural network (CNN) backbone with a Feature Pyramid Network (FPN) module, an encoder block, a decoder block with a detection branch and a segmentation branch, followed by an optional SAM-Med2d block.

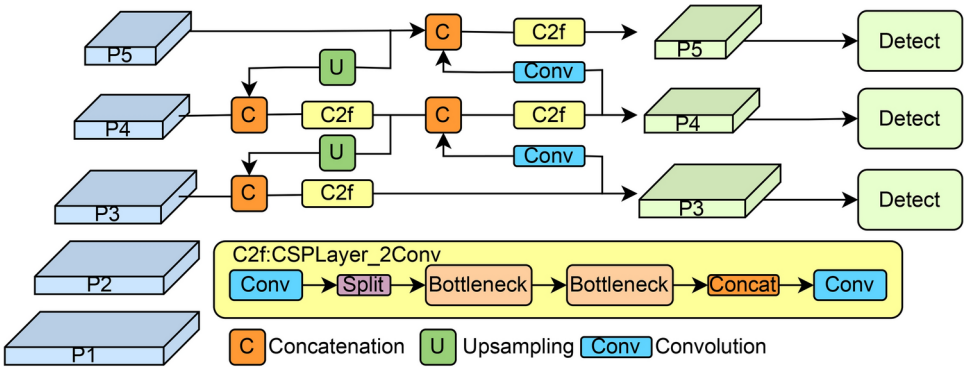

Fig. 2. Detection branch architecture. The P1 to P5 blocks (in blue) represent different levels of the feature pyramid, each corresponding to a distinct scale of feature representation. This enables the model to simultaneously detect targets of varying sizes. P3 to P5 in green are blocks in the decoder which also represent different levels of the feature pyramid, responsible for the final classification and bounding box regression of the extracted features.

the challenge of limited pixel-level annotations by leveraging weakly supervised learning techniques, thereby reducing the need for extensive manual labeling. The overall framework of WSFS, as shown in Fig. 1, integrates multiple components to achieve high-precision segmentation. It consists of an encoder block, a decoder with a detection branch and a segmentation branch, followed by an optional SAM-Med2D block.

WSFS employs CSPDarknet[31] as the backbone/encoder for feature extraction, handled by a series of convolutional and deconvolutional layers, complemented by residual connections[16] and bottleneck structures[50] to reduce the size of the network and enhance performance. The C2f module is used as the fundamental building block to improve computational efficiency. The backbone is augmented with a Feature Pyramid Network (FPN)[51] module, which facilitates multi-scale feature representation, crucial for capturing the varying sizes and shapes of follicles. The detection branch operates on multiple feature maps generated by the FPN and predicts bounding boxes and class probabilities for potential follicles. And the segmentation branch is the core component of WSFS, leveraging the concept of MIL for weakly supervised segmentation.

Detection branch loss

The network's detection branch is purpose-built to pinpoint probable follicle areas and generate bounding boxes around the targets by leveraging a CNN foundation. WSFS draws inspiration from the YOLOv8[31] method, a state-of-the-art object detection algorithm known for its efficiency and accuracy. As shown in Fig. 2, this branch operates on feature maps generated by the FPN (specifically P3, P4, and P5 levels), predicting bounding boxes and class probabilities for potential follicles. yolov8x-seg adopts the C2f module within its core CSPDarkNet backbone and subsequent neck component. The network employs a meticulously crafted loss function that is geared towards refining its detection precision, which can be expressed as:

$$\mathcal{L}_{\text{det}} = \mathcal{L}_{\text{Bbox}} + \mathcal{L}_{\text{Cls}}$$

Where $\mathcal{L}_{\text{Bbox}}$ is the bounding box regression loss, leveraging Complete Intersection over Union Loss (CIoU) loss and Distribution Focal loss, to measure the regression degree of the bounding box. In conjunction with the Anchor-Free approach and improving generalization, it allows the network to quickly focus on the distribution of the target position and its surrounding areas. And the $\mathcal{L}_{\text{Cls}}$ is the classification loss, leveraging Binary Cross-Entropy (BCE) Loss, a common loss function used in binary classification tasks, to classify whether the target is a follicle. After prediction, Non-Maximum Suppression (NMS) is applied to eliminate redundant bounding boxes, producing a refined set of detection.

Mask branch MIL loss

In this work, we employ MIL framework[13] as a core component for weakly supervised instance segmentation by leveraging the bounding box tightness prior to handle the coarse annotations. The tightness prior generates positive and negative bags based on the sweeping lines within each bounding box (Fig. 3). A positive bag is composed of pixels that intersect the bounding box, and thus must contain at least one pixel of the object, while a negative bag consists of pixels that do not intersect any bounding boxes, indicating no presence of the object. Denote b_i the set of mask probabilities of the pixel instances belonging to bag i, the MIL loss is defined:

$$\mathcal{L}_{\text{mil}} = -\sum_i y_i \log(\max b_i) - (1 - y_i) \log(1 - \max b_i)$$

where $y_i = 1$ if bag i is positive, and $y_i = 0$ otherwise.

The multi-task loss

The overall loss function combines the detection and segmentation losses to perform multi-task learning. This encourages the network to learn features that are beneficial both for detection and segmentation tasks. The loss of multiple tasks is given by:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{det}} + \alpha \mathcal{L}_{\text{mil}}$$

Where α is a hyperparameter that balances the contribution of the MIL loss relative to the detection loss.

The main advantage of our approach is its ability to leverage the tightness prior to bounding boxes to improve the accuracy of instance segmentation under weak supervision. By integrating MIL with bounding box constraints, our method effectively reduces the annotation burden while enhancing segmentation precision. It addresses challenges such as handling variations in bounding box tightness and ensuring robustness against noisy annotations.

Plug and play to SAM-Med2D

SAM and its adaptation for medical images, SAM-Med2D[15] can yield better performance with valuable prompts. Given the light weighted nature of the WSFS, it is designed to integrate seamlessly with SAM-Med2D, which is a state-of-the-art framework for medical image segmentation. Namely, the outputs generated by WSFS, whether in the form of point, bounding boxes, or masks, can be used as prompts for SAM-Med2D. This integration leverages the preliminary segmentation results from WSFS and the general capabilities of a pre-trained large vision model to refine and enhance the overall accuracy of follicle segmentation in ultrasound images.

Ethical statement

This study was reviewed and approved by the Ethical Committee of Xiangya Hospital, Central South University (Ethical Approval Number: 2023005). All experiments were conducted in accordance with the ethical standards outlined in the Declaration of Helsinki and followed relevant national and international guidelines and regulations. Informed consent was obtained from all participants and/or their legal guardians prior to their involvement in the study. Confidentiality and anonymity of all participants were strictly maintained throughout the research process. The study design and procedures were carefully reviewed to ensure compliance with ethical principles, minimizing risks to participants while upholding the integrity of the research.

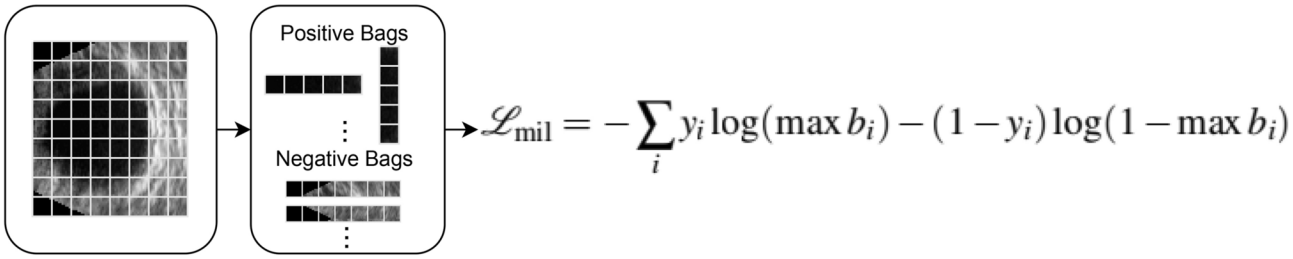

Fig. 3. MIL loss function. It enforces the tightness constraints of the bounding boxes on the prediction map. It encourages the model to predict a high score for at least one pixel in a positive bag (a region within a bounding box that should contain the object) and a low score for all pixels in a negative bag (a region outside any bounding box that should not contain the object). b_i represents the set of mask probabilities of the pixel instances belonging to bag i, $y_i = 1$ if bag i is positive, and $y_i = 0$ otherwise.

Experiment and results

Dataset and evaluation metrics

- *Dataset.* For evaluation, we utilized two distinct datasets in our experiments: the publicly available USOVA3D dataset[52] and our FUID dataset. The USOVA3D dataset, which is specifically curated for the analysis of follicles and ovarian ultrasound images, was employed for training and validating our model. To align with the input dimensions required by our model, the 3D follicle data from USOVA3D was processed by slicing along the z-axis, thereby converting the volumetric data into a series of 2D cross-sectional images. The USOVA3D public database was established specifically to validate the effectiveness of follicle detection algorithms on 3D ovarian ultrasound images. The USOVA3D database comprises 35 volumes of ovarian ultrasound data that have been annotated by two medical experts. These data are processed and saved in VTK format using the ITK-SNAP[53] tool to facilitate 3D image analysis. In addition to the USOVA3D dataset, This study has successfully developed a novel follicle ultrasound image dataset, Follicle Ultrasound Image Dataset (FUID), which is a significant contribution to the community. Recognizing the scarcity of such resources in the domain of reproductive health and computer-aided diagnosis, this dataset aims to fill the gap and provide a rich source of data for future studies. FUID, which comprises a substantial number of ultrasound images, was instrumental in evaluating how well our model could adapt to and perform on data that were not seen during the training phase. By including FUID in our experimental validation, we aimed to ensure that our model's performance demonstrates its robustness and applicability across different and real clinical datasets. This approach is crucial for confirming the generalization of our model in real-world scenarios. The existing USOVA3D dataset, which is the only publicly available counterpart, offers a limited view with just 35 follicle and ovarian ultrasound image cases. In comparison, the FUID in this study expands the available data exponentially, encompassing 193 distinct cases with a total of 22,942 high-quality ultrasound images. This extensive collection is designed to cater to the needs of various research endeavors, including but not limited to, the development and validation of algorithms for automatic follicle detection, characterization, and monitoring. To safeguard patient privacy and adhere to strict ethical guidelines, all images within the dataset have been meticulously processed to remove any identifiable information. This de-identification ensures that the dataset can be freely used by researchers worldwide without compromising the privacy rights of the individuals depicted in the images. On FUID, the pixel-wise semantic annotations are provided by 2 experts of Assisted Reproduction department. Each image is first annotated by one expert and then double checked by another, which guarantees the annotating quality. Informed consent was obtained for experiments in this study.
- *Evaluation metrics.* The standard evaluation metrics, the mean average precision (mAP), Intersection over Union (IOU) and Dice Score are adopted. mAP50 is a performance metric used to evaluate the quality of an object detection system across all classes (in our case only one class as follicle) and is particularly focused on the precision at a 50% IOU threshold. It is computed by first calculating the Average Precision (AP) at the 50% IOU threshold. AP is the area under the precision-recall curve , where precision is defined as the number of true positives divided by the total number of predicted positives (true positives plus false positives) and recall is the number of true positives divided by the total number of actual positives. The formula for AP is:

$$\mathrm{AP} = \sum_{n=1}^{N} (\mathrm{Precision}_n - \mathrm{Precision}_{n-1}) \times \mathrm{Recall}_n$$

mAP50 is then the mean of the AP scores across all classes, giving a single number that summarizes the performance of the model in different object categories. The formula for mAP50 is:

$$\mathrm{mAP}_{50} = \frac{1}{C} \sum_{c=1}^{C} \mathrm{AP}_c$$

where C is the total number of classes and is the AP for class c. The IOU is a metric used to evaluate the accuracy of predictions by measuring the overlap between the predicted bounding box and the GT bounding box. It is calculated as the ratio of the overlap area between the two bounding boxes to the overlap area of their union. The formula of IOU is:

$$\mathrm{IOU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

The IOU score ranges from 0 to 1, where 0 indicates no overlap and 1 indicates perfect alignment. Similarly, the Dice Score is a metric used to measure the overlap between two sets A and B, commonly applied in image segmentation tasks to evaluate the accuracy of predicted segmentation against the GT. The formula of Dice Score is :

$$\text{Dice Score} = \frac{2|A \cap B|}{|A| + |B|}$$

where A represents the prediction set and B represents the GT set. We report mAP50, IOU and Dice Score in the USOVA3D dataset.

Method	mAP50	IOU	Dice Score	FLOPs(G)	Params(M)
WSFS	0.911	0.682	0.81	10.5	71.8
WSFS+	0.957	0.714	0.83	10.5	71.8

Table 1. Quantitative comparisons of WSFS and WSFS+ on USOVA3D dataset.

Method	mAP50	IOU	Dice Score	FLOPs(G)	Params(M)
U-Net*	0.656	0.612	0.76	21.0	*31.3*
ADGC-UNet*	–	0.612	0.76	*11.2*	**6.3**
Yolov8x-seg*	*0.920*	**0.789**	**0.88**	344.1	71.8
DiscoBox	0.747	0.346	0.51	71.1	44.8
SAM-Med2D	0.716	0.373	0.54	33.8	271.2
WSFS+	**0.957**	*0.714*	*0.83*	**10.5**	71.8

Table 2. Quantitative comparisons of WSFS+ with the state-of-the-art methods including U-Net, ADGC-UNet, Yolov8x-seg, DiscoBox, and SAM-Med2D on USOVA3D dataset. The best and second-best results in each category are highlighted in bold and italic formats, respectively. * represents fully supervised method.

Implementation details

We implement the proposed method using PyTorch. During training, the network is optimized on a machine with two GeForce GTX 3090 GPUs. The batch size, learning rate, weight decay, momentum, and the number of iterations are set to 128, $3 * 10^{-2}$, $2 * 10^{-4}$, 0.95 and 22k, respectively. We choose stochastic gradient descent (SGD)[54] as the optimization solver. For data augmentation, following the setting used in yolov8x-seg, we horizontally and vertically flip each image with probability 0.5, and set the image size to 224 as the USOVA3D slices have image size not exceeding 212 pixels.

Experiment results

We initially embarked on training the WSFS model using public datasets USOVA3D, aiming to establish a performance baseline. As shown in Table 1, WSFS achieved performance with mAP50 of 0.911, IOU of 0.682 and Dice Score of 0.81. The results obtained from WSFS demonstrated a competitive performance compared to other state-of-the-art methods presented in Table 2, demonstrating the robustness of our method even without the inclusion of additional private data. However, we encountered limitations due to the inherent constraints of the data-namely, the relatively small sample size and the high noise levels. Recognizing the need for a more robust training regimen, we integrate private FUID data into our training process. This strategic augmentation of our dataset led to a marked improvement in the model's performance, resulting in the enhanced WSFS+ model with mAP50 of 0.957, IOU of 0.714 and Dice Score of 0.83. The enhanced WSFS+ model showed a significant increase in accuracy and reliability, highlighting the effectiveness of our proposed method. Our findings suggest that with the incorporation of more high-quality ultrasound image data, the WSFS+ model has the potential for further refinement, ultimately leading to even more precise and dependable segmentation capabilities.

Next, using WSFS+, we compared its performance against State-of-the-Art methods. As shown in Table 2, our proposed method, WSFS+, demonstrates a significant performance improvement over existing state-of-the-art methods when evaluated on the USOVA3D dataset. With a mAP50 of 0.957, IOU of 0.714 and Dice Score of 0.83, WSFS+ outperforms the classic fully supervised method U-Net[10], which achieved a mAP50 of 0.656, IOU of 0.612 and Dice Score of 0.76, and the latest proposed variant ADGC-UNet[22] with a performance of IOU of 0.612 and Dice Score of 0.76. This demonstrates that our weakly supervised approach offers promising segmentation performance compared to fully supervised methods. Notably, even the advanced YOLOv8x-seg[31], which inspired the detection branch of WSFS+, shows a lower mAP50 of 0.920. Although YOLOv8x-seg performs better in terms of IOU due to the dense information learned from pixel-level annotations, it requires significantly higher computational resources in terms of FLOPs (Floating Point Operations). Upon comparative analysis with the DiscoBox model, the state-of-the-art weakly supervised learning method, our WSFS+ model exhibited a significantly superior performance. The qualitative assessment (Fig. 5) revealed that DiscoBox had difficulty with follicle segmentation. Similarly, the SAM-Med2D model, despite being specifically designed and pre-trained on numerous medical image datasets, did not meet performance expectations. Given its robust training on diverse medical imagery, SAM-Med2D seemed to falter when applied to ultrasound images, suggesting that even specialized pre-training may not guarantee optimal results across all medical imaging scenarios. This observation further supports the notion that instance segmentation in the context of ultrasound images presents a formidable challenge, primarily due to the unique characteristics and complexities of these images. We also conjecture that the overall performance of the SAM serial method heavily depends on the quality of the prompts.

We compared the performance of SAM-Med2D with our method, as shown in Table 3 and Fig. 6, WSFS+ yielded more impressive results. We found that despite SAM-Med2D's extensive pre-training, it struggled with the segmentation of follicle ultrasound images, achieving a mAP50 of 0.716, IOU of 0.373 and Dice Score of 0.54. This difficulty is primarily due to the absence of relevant ultrasound data in its pre-training corpus and the

inherently noisy nature of ultrasound imagery. This limitation is also common in zero-shot learning within the SAM series pipeline.

To further enhance SAM-Med2D's performance, we utilized the outputs from WSFS, including points, bounding boxes, and masks, as prompts for SAM-Med2D. This strategy leveraged the fine-grained information provided by WSFS to guide the segmentation capabilities of SAM-Med2D, resulting in a noticeable improvement in segmentation accuracy. The combined approach represented as WSFS+(mask) &SAM-Med2D, which leveraged the output of WSFS+ as the mask prompt for SAM-Med2D, attains a mAP50 of 0.967, IOU of 0.724 and Dice Score of 0.84, while maintaining a reasonable computational cost with 44.3G FLOPs and 343M parameters. This synergistic effect of WSFS+ and SAM-Med2D illustrates the potential of our method when augmented with advanced post-processing techniques, solidifying its position as a leading solution in follicle segmentation for ultrasound imaging.

The qualitative results from this combined approach (SAM-Med2D with WSFS+ mask prompt) were promising, illustrating that even a large pre-trained model like SAM-Med2D could benefit substantially from the finely tuned segmentation cues provided by WSFS+. This synergistic integration not only improved the segmentation capabilities of SAM-Med2D but also validated the effectiveness of our WSFS+ as a powerful prompt for enhancing the performance of larger models.

The results demonstrated that our proposed model, WSFS+, when trained from scratch with a combination of public and private ultrasound image data, could achieve competitive results that surpassed those of models pre-trained on more general datasets. Moreover, the qualitative analysis highlighted the potential of WSFS+ as an effective prompt generator for large models such as SAM-Med2D, particularly in the challenging domain of ultrasound image segmentation. This study demonstrates the importance of tailored training data and the integration of weakly supervised learning with sophisticated post-processing frameworks in advancing the field of medical image analysis.

Figure 4 illustrates examples of follicle instance segmentation using our WSFS and enhanced WSFS+ methods. Comparisons of these segmentations with other methods are shown in Fig. 5. Figure 6 demonstrates the SAM-Med2D with different prompts generated by WSFS and enhanced WSFS+. Our proposed method can effectively learn useful information from high-quality data and further optimize performance as the training dataset size increases. The results of Fig. 5 show that in fully supervised learning methods, yolov8x-seg outperforms U-net in all aspects, due to its more advanced architecture. yolov8x-seg also has the best IOU performance.

The weakly supervised learning method, DiscoBox, exhibits relatively poor segmentation performance compared to other methods, see Fig. 5. We can observe from Fig. 6 - row SAM-Med2D that the state-of-the-art method also falls short of expectations, with predictions showing numerous fragments and gaps, indicating it cannot accurately identify follicles. WSFS+ achieved the best mAP50 result, outperforming other methods in accurate recognition and clear boundary convergence.

Notably, Fig. 6 results indicate that, based on SAM-Med2D, the point prompts generated by both WSFS and WSFS+ have little impact on its prediction results, even though the points correctly located the follicles. We conjecture that this is due to the noise in ultrasound images reducing the homogeneity of objects, resulting in fragmented segmentation. Compared to point prompts, box prompts significantly improve predictive performance, but gaps and overlap issues remain. WSFS+(mask) &SAM-Med2D has the best performance in mAP50, IOU and Dice Score metrics, and can accurately distinguish adjacent follicles, showing the best results.

Discussion

The WSFS+ model demonstrated a remarkable improvement in performance over its predecessor, demonstrating a significant leap in the accuracy of follicle segmentation within ultrasound images. This improvement can be attributed to the proposed simple yet efficient weakly supervised learning architecture and the strategic integration of private data into the training set, resulting in a richer and more diverse dataset. The model's architecture, fine-tuned for the ultrasound imagery, proved adept at capturing the subtleties required for precise segmentation. The qualitative and quantitative results indicate that WSFS+ is a robust solution for follicle segmentation in ultrasound images, competitive with fully supervised methods.

Method	mAP50	IOU	Dice Score	FLOPs(G)	Params(M)
WSFS+	*0.957*	*0.714*	*0.83*	**10.5**	**71.8**
SAM-Med2D	0.716	0.373	0.54	*33.8*	*271.2*
WSFS(pt) & SAM-Med2D	0.716	0.590	0.74	44.3	343.0
WSFS(box) & SAM-Med2D	0.938	0.700	0.82	44.3	343.0
WSFS(mask) & SAM-Med2D	0.921	0.703	0.83	44.3	343.0
WSFS+(pt) & SAM-Med2D	0.795	0.596	0.75	44.3	343.0
WSFS+(box) & SAM-Med2D	0.942	0.707	0.83	44.3	343.0
WSFS+(mask) & SAM-Med2D	**0.967**	**0.724**	**0.84**	44.3	343.0

Table 3. Quantitative comparisons of WSFS+ with SAM-Med2D method and the combinations of WSFS/WSFS+ with SAM-Med2D where SAM-Med2D leverages the output of WSFS/WSFS+ in forms of point/bounding box/mask as prompt on USOVA3D dataset. The best and second-best results in each category are highlighted in bold and italic formats.

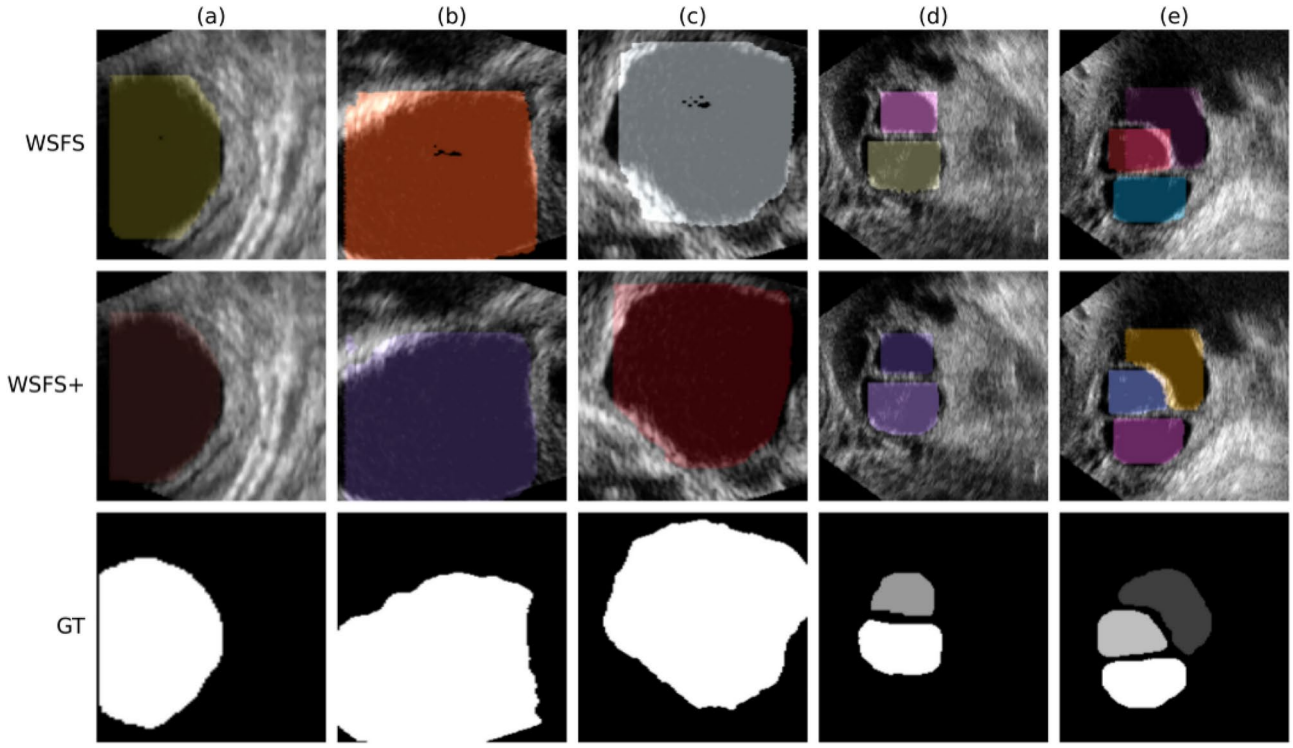

Fig. 4. Examples of follicle instance segmentation by our WSFS and enhanced WSFS+.

Despite the prevalence of fully supervised methods in state-of-the-art methods, the unique challenges posed by ultrasound images necessitate more specialized approaches. The SAM-Med2D model, while pre-trained on a variety of medical datasets, did not perform satisfactorily when directly applied to ultrasound images. This under-performance suggests that the generalization capabilities of pre-trained models may be limited when dealing with the specific characteristics of ultrasound data. The high noise levels and the need for detailed anatomical understanding appear to have hindered the model's ability to segment images effectively.

Within the framework of WSFS+, SAM-Med2D was repurposed to function as an output refinement module. This strategic shift leveraged SAM-Med2D's strengths in post-processing, enhancing the final output of the WSFS+ model. By focusing on refining the segmentation masks generated by WSFS+, SAM-Med2D contributed to a cleaner and more accurate final product. Meanwhile, WSFS+ can also be considered a prompt generator to fully leverage the potential of SAM-Med2D. This dual approach highlights the complementary nature of the two models and underscores the importance of a multi-faceted strategy in addressing complex medical imaging tasks. Moreover, this study demonstrates that, instead of adopting downstream fine-tuning based on large pre-trained models, generating optimized prompts is a more efficient alternative to explore the potential of pre-trained models.

The pursuit of enhanced performance in medical image segmentation often comes with a trade-off with respect to computational resources. While the WSFS+ model offers superior accuracy, it also has significantly lower demand on computational resources in terms of FLOPs and parameters, which makes WSFS+ readily deployable in real complicated clinical scenarios where computational resources are limited. On the other hand, the combination of WSFS+ and SAM-Med2D can be a good option to further optimize the prediction results, yet may require more substantial computational power and memory during training and inference. This trade-off must be carefully considered, particularly in clinical settings where both high performance and computational efficiency are paramount.

Our experimental results demonstrate the effectiveness of the proposed weakly supervised instance segmentation method for ultrasound images of the follicle. Despite the use of coarse annotations, our model achieved competitive segmentation performance compared to fully supervised approaches. By leveraging coarse annotations in the form of bounding boxes, our method offers a practical solution for automating follicle segmentation in clinical settings. This advancement has the potential to streamline follicle development monitoring processes, providing clinicians with efficient tools for more efficient and effective fertility treatment.

The proposed method in this study represents a significant advancement in the field of weakly supervised medical image segmentation. It demonstrates the potential of leveraging weak annotations to achieve segmentation performance on par with fully supervised approaches, opening up new possibilities for research and clinical applications where obtaining extensive annotations is impractical or costly. However, our method has certain limitations that need to be addressed in the future. These can be summarized as follows: Firstly, the current structural design could be further simplified to enhance efficiency and facilitate easier integration with other systems. Specifically, the absence of end-to-end integration with SAM limits seamless data flow

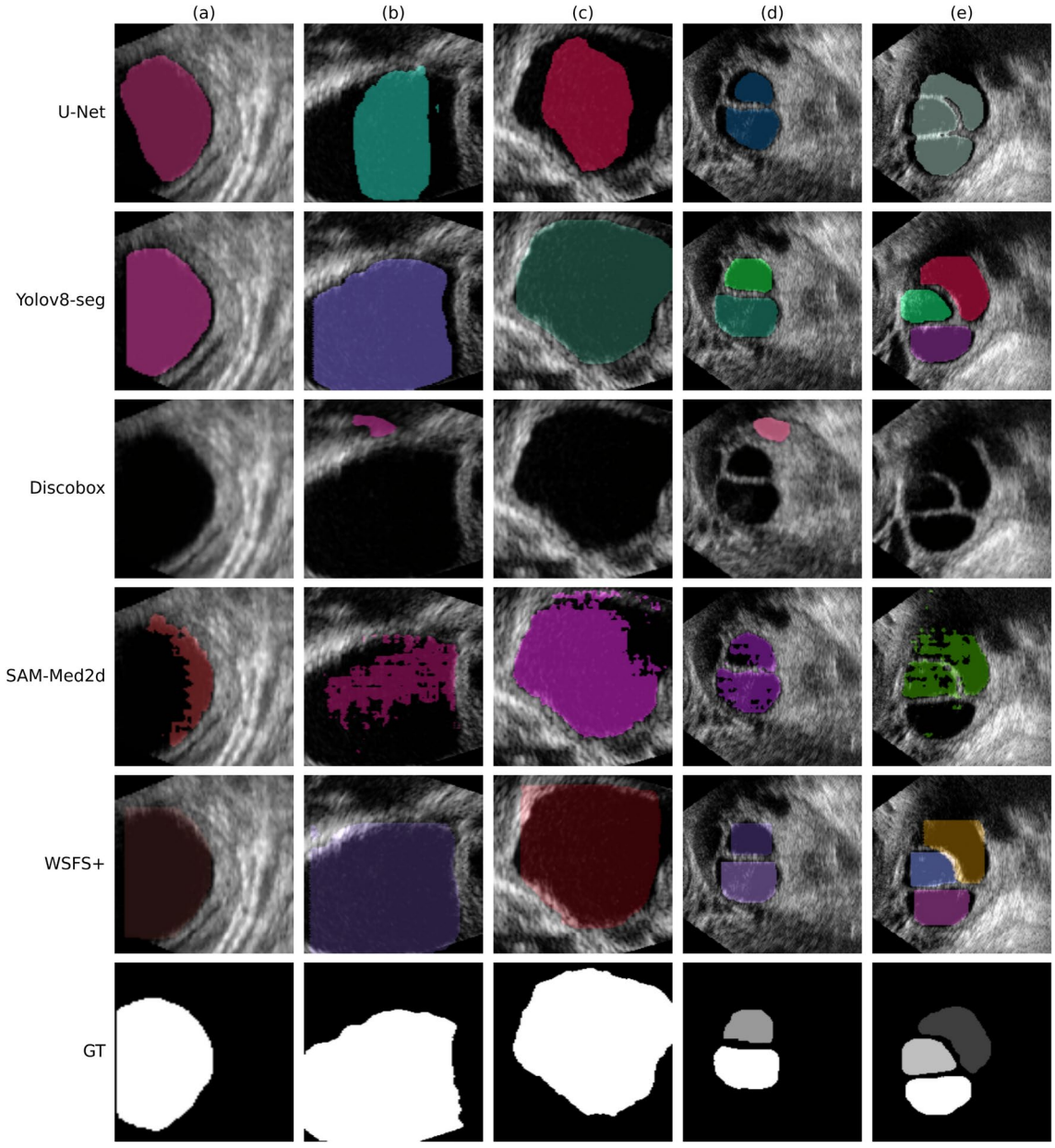

Fig. 5. Examples of follicle instance segmentation by our WSFS+, U-Net, YOLOv8x-seg, DiscoBox, and SAM-Med2D.

and interaction, which could otherwise improve overall system performance. Secondly, the method is not yet optimized for high-resolution scenarios, as it currently supports images of size 224. Introducing a feature to dynamically handle high-resolution images as input is a potential enhancement. Thirdly, the approach requires improvements in handling extremely noisy ultrasound images. While addressing such cases remains an open challenge, enhancing the robustness of the method in these scenarios is crucial for evaluating its adaptability and effectiveness across diverse conditions. In addition, our method can also be applied to the segmentation task of other structures, only requiring bounding-box-level annotations to accomplish the task of dense prediction. We will further explore and expand the application scope of our method in future work.

Conclusion

Follicle segmentation in ultrasound images is crucial for monitoring follicle development in fertility treatments. However, it remains challenging due to the lack of ground truths (GT) and available datasets, particularly in supervised learning. This study introduces a new dataset and proposes a weakly supervised method to address this issue. The proposed Weakly Supervised Follicle Segmentation (WSFS) method is a novel one-stage segmentation pipeline that uses bounding boxes as approximate labels and employs Multiple Instance Learning (MIL) for weakly supervised learning. The WSFS method has shown promising performance, achieving

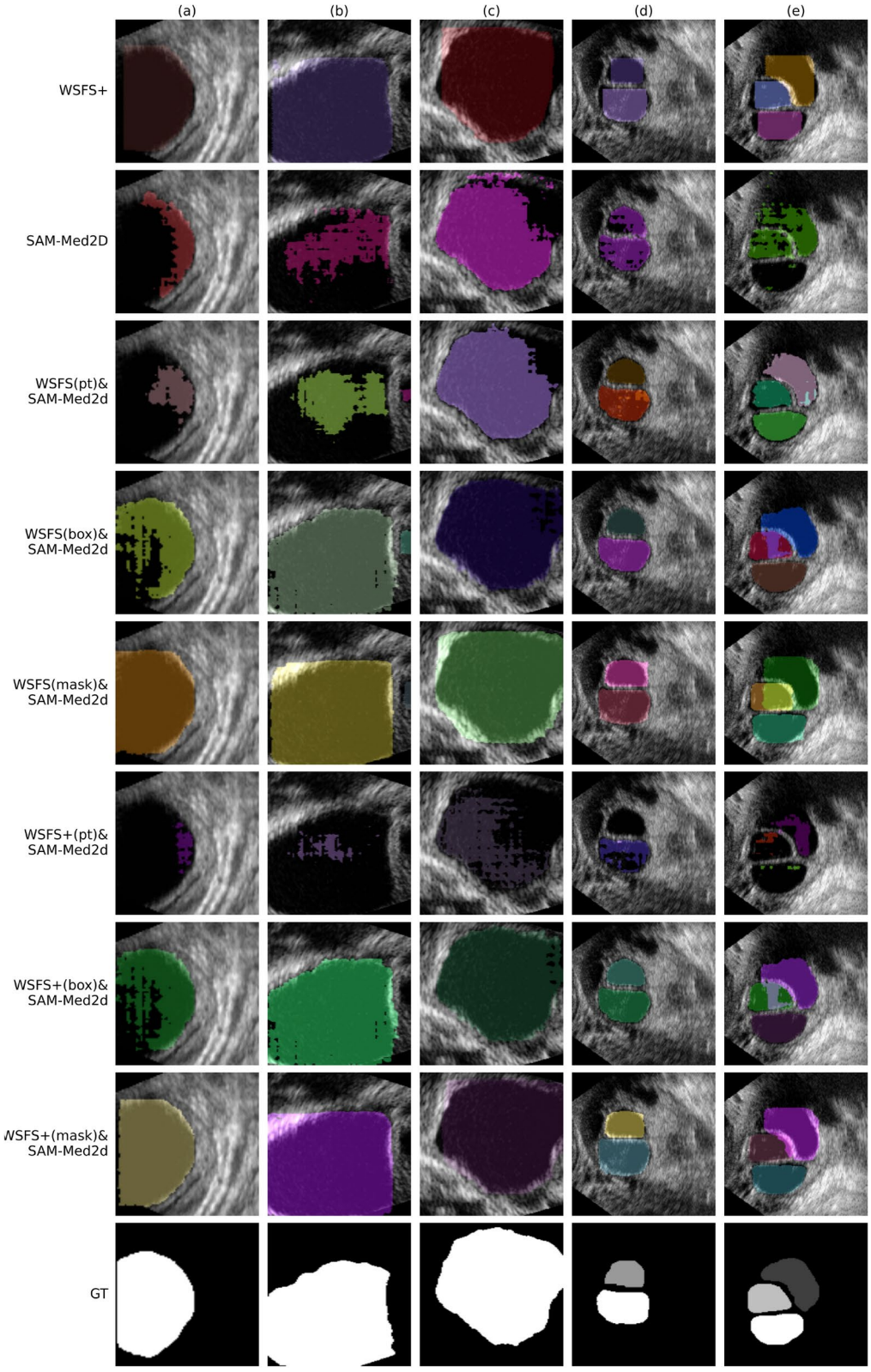

Fig. 6. Examples of follicle instance segmentation by our WSFS+, SAM-Med2D, and SAM-Med2D with point/bounding box/mask prompt from WSFS/WSFS+.

competitive results with fully supervised methods. The WSFS+ model achieves a mAP50 of 0.957, IOU of 0.714 and Dice Score of 0.83. Additionally, when used as a prompt generator within SAM-Med, the WSFS&SAM-Med model achieves state-of-the-art performance with a mAP50 of 0.967, IOU of 0.724 and Dice Score of 0.84. This dual approach highlights the complementary nature of the two models and underscores the importance of a multifaceted strategy in tackling complex medical imaging tasks. Furthermore, we demonstrated a new paradigm where generating optimized prompts is an efficient alternative to downstream fine-tuning based on large pre-trained models, exploring the potential of pre-trained models more effectively.

Data availability

The data that support the findings of this research is collected from (1) USOVA3D dataset https://doi.org/10.1016/j.cmpb.2020.105621 (2) FUID dataset, which was available on the Follicle-ultrasound-image-dataset-FUID-repository. https://github.com/charliecaviar/Follicle-ultrasound-image-dataset-FUID. The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Received: 23 July 2024; Accepted: 25 March 2025
Published online: 21 April 2025

References

1. Farquhar, C., Rishworth, J. R., Brown, J., Nelen, W. L. D. M. & Marjoribanks, J. Assisted reproductive technology: An overview of cochrane reviews. *Cochrane Datab. Syst. Rev.* CD010537 (2015).
2. Racca, A., Drakopoulos, P., Neves, A. R. & Polyzos, N. P. Current therapeutic options for controlled ovarian stimulation in assisted reproductive technology. *Drugs* **80**, 973–994 (2020).
3. Azad, R. *et al.* Medical image segmentation review: The success of u-net. arxiv.org/abs/2211.14830v1 (2022).
4. Ma, J. et al. Segment anything in medical images. *Nat. Commun.* **15**, 654 (2024).
5. Yao, W. et al. *From cnn to transformer: A review of medical image segmentation models*. arXiv:2308.05305 (2023).
6. Yadav, N. Despeckling filters applied to thyroid ultrasound images: A comparative analysis. *Multimedia Tools and Applications* (2022).
7. Yadav, N., Dass, R. & Virmani, J. A systematic review of machine learning based thyroid tumor characterisation using ultrasonographic images. *J. Ultrasound* **27**, 209–224 (2024).
8. Yadav, N. Objective assessment of segmentation models for thyroid ultrasound images. *J. Ultrasound* (2023).
9. Wang, H. et al. Application of deep convolutional neural networks for discriminating benign, borderline, and malignant serous ovarian tumors from ultrasound images. *Front. Oncol.* **11**, 770683 (2021).
10. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. arXiv:1505.04597 (2015).
11. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3d u-net: Learning dense volumetric segmentation from sparse annotation. arXiv:1606.06650 (2016).
12. He, K., Gkioxari, G., Dollár, P. & Girshick, R. *Mask r-cnn*. arXiv:1703.06870 (2018).
13. Hsu, C.-C., Hsu, K.-J., Tsai, C.-C., Lin, Y.-Y. & Chuang, Y.-Y. Weakly supervised instance segmentation using the bounding box tightness prior.
14. Kirillov, A. et al. *Segment anything*. arXiv:2304.02643 (2023).
15. Cheng, J. *et al.* Sam-med2d. arXiv:2308.16184 (2023).
16. He, K., Zhang, X., Ren, S. & Sun, J. *Deep residual learning for image recognition*. arXiv:1512.03385 (2015).
17. Li, H. et al. Cr-unet: A composite network for ovary and follicle segmentation in ultrasound images. *IEEE J. Biomed. Health Inform.* **24**, 974–983 (2020).
18. Singh, V. K. et al. Hatu-net: Harmonic attention network for automated ovarian ultrasound quantification in assisted pregnancy. *Diagnostics* **12**, 3213 (2022).
19. Christiansen, F. et al. Ultrasound image analysis using deep neural networks for discriminating between benign and malignant ovarian tumors: Comparison with expert subjective assessment. *Ultrasound Obstet. Gynecol.* **57**, 155–163 (2021).
20. Zhao, Q. *et al.* Mmotu: A multi-modality ovarian tumor ultrasound image dataset for unsupervised cross-domain semantic segmentation. arXiv:2207.06799 (2023).
21. 3d convolutional neural networks for human action recognition | ieee journals & magazine | ieee xplore. https://ieeexplore.ieee.org/document/6165309.
22. Sarkar, M. & Mandal, A. Attention gated double contraction path u-net for follicle segmentation from ovarian usg images. *Multimedia Tools and Applications* .
23. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788 (IEEE, Las Vegas, NV, USA, 2016).
24. Redmon, J. & Farhadi, A. Yolo9000: Better, faster, stronger. arXiv:1612.08242 (2016).
25. Redmon, J. & Farhadi, A. Yolov3: An incremental improvement. arXiv:1804.02767 (2018).
26. Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y. M. Yolov4: Optimal speed and accuracy of object detection. arXiv:2004.10934 (2020).
27. Ultralytics/yolov5: Yolov5 in pytorch pytorch > onnx > coreml > tflite https://github.com/ultralytics/yolov5.
28. Ge, Z., Liu, S., Wang, F., Li, Z. & Sun, J. Yolox: Exceeding yolo series in 2021. arxiv.org/abs/2107.08430v2 (2021).
29. Li, C. *et al.* Yolov6: A single-stage object detection framework for industrial applications. arXiv:2209.02976 (2022).
30. Wang, C.-Y., Bochkovskiy, A. & Liao, H.-Y. M. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv:2207.02696 (2022).
31. Ultralytics/ultralytics: New - yolov8 in pytorch > onnx > openvino > coreml > tflite. https://github.com/ultralytics/ultralytics.
32. Wang, C.-Y., Yeh, I.-H. & Liao, H.-Y. M. Yolov9: Learning what you want to learn using programmable gradient information. arXiv:2402.13616 (2024).
33. Krähenbühl, P. & Koltun, V. *Efficient inference in fully connected crfs with gaussian edge potentials*. arXiv:1210.5644 (2012).
34. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. *Semantic image segmentation with deep convolutional nets and fully connected crfs*. arXiv:1412.7062 (2016).
35. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv:1606.00915 (2017).
36. Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. *Rethinking atrous convolution for semantic image segmentation*. arxiv:1706.05587 (2017).
37. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. arxiv.org/abs/1802.02611v3 (2018).
38. Cheng, B., Schwing, A. G. & Kirillov, A. *Per-pixel classification is not all you need for semantic segmentation*. arxiv:2107.06278 (2021).

39. Ren, S., He, K., Girshick, R. & Sun, J. *Faster r-cnn: Towards real-time object detection with region proposal networks*. arXiv:1506.01497 (2016).
40. Liu, S., Qi, L., Qin, H., Shi, J. & Jia, J. *Path aggregation network for instance segmentation.* **1803**, 01534 (2018).
41. Wang, X., Kong, T., Shen, C., Jiang, Y. & Li, L. Solo: Segmenting objects by locations. arXiv:1912.04488 (2020).
42. Wang, X., Zhang, R., Kong, T., Li, L. & Shen, C. Solov2: Dynamic and fast instance segmentation. arXiv:2003.10152 (2020).
43. Bolya, D., Zhou, C., Xiao, F. & Lee, Y. *J. Yolact: Real-time instance segmentation* **1904**, 02689 (2019).
44. Bolya, D., Zhou, C., Xiao, F. & Lee, Y. J. Yolact++: Better real-time instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 1108–1121 (2022).
45. Li, W. *et al.* Box2mask: Box-supervised instance segmentation via level-set evolution. arXiv:2212.01579 (2022).
46. Rother, C., Kolmogorov, V. & Blake, A. "grabcut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **23**, 309–314 (2004).
47. Pont-Tuset, J., Arbelaez, P., Barron, J. T., Marques, F. & Malik, J. Multiscale combinatorial grouping for image segmentation and object proposal generation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 128–140 (2017).
48. Tian, Z., Shen, C., Wang, X. & Chen, H. Boxinst: High-performance instance segmentation with box annotations. arXiv:2012.02310 (2020).
49. Lan, S. *et al.* Discobox: Weakly supervised instance segmentation and semantic correspondence from box supervision. Arxiv:2105.06464 (2021).
50. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. arXiv:1801.04381 (2019).
51. Lin, T.-Y. et al. *Feature pyramid networks for object detection*. arXiv:1612.03144 (2017).
52. Potočnik, B. et al. Public database for validation of follicle detection algorithms on 3d ultrasound images of ovaries. *Comput. Methods Programs Biomed.* **196**, 105621 (2020).
53. Yushkevich, P. A. et al. User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* **31**, 1116–1128 (2006).
54. Bottou, L. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010* (eds Lechevallier, Y. & Saporta, G.) 177–186 (Physica-Verlag HD, 2010).

Acknowledgements

This study was funded by the National Natural Science Foundation of China (grant no. 82371682), the National Key Research and Development Program of China (grant no. 2021YFC2700404; 2022YFC2010200), the Natural Science Foundation of Hunan Province (grant no. 2022JJ40779; 2022JJ70080), the Fundamental Research Funds for the Central Universities of Central South University (grant no. 1053320222496), and the Science and Technology Innovation Program of Hunan Province (grant no. 2022RC3013).

Author contributions

G.L. led the conceptualization and design of the study and the experiments, developed the methodology, and contributed to the writing of the manuscript, L.Z. contributed the design, analysis and revision of the manuscript, Z.Z. participated in the experiments implementation, and contributed to the data analysis, J.H. assisted in data curation and pre-processing, H.T. and J.F. involved in the collection of the medical image dataset and contributed to the validation of the model, Q.Z. provided administrative support and helped in coordinating the research activities, W.H. and Y.L. provided supervision and mentorship throughout the research process, Y.W. and J.H. provided overall guidance for the project, supervised the research, and contributed to the finalization of the manuscript. All authors reviewed the manuscript.

Declarations

Competing interests

All authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.W. or J.H.