YOSAM: A YOLO and MedSAM-Based Framework for Automatic Measurement of Fetal Head Circumference in Ultrasound Images

Zhihao Li[1], Liyan Chen[1], Li Lu[1], Ye Ding[1(✉)], and Xiuxiu Hao[2]

[1] School of Computer Science and Technology, Dongguan University of Technology, Dongguan 523808, China
{241115318,241115300,221115220,dingye}@dgut.edu.cn
[2] The Fifth Medical Center of PLA General Hospital, Beijing, China

Abstract. Accurate measurement of fetal head circumference (HC) in ultrasound images remains essential yet challenging for obstetric assessment, primarily due to anatomical variations across gestational stages and inherent imaging artifacts. In response to these limitations, we introduce YOSAM, a novel framework for fetal HC measurement that synergistically combines YOLOv11-based detection with our enhanced MedSAM-AD model. The MedSAM-AD integrates an Adapter layer for domain-specific feature adaptation and a Dimensional Reciprocal Attention Mixing Transformer (D-RAMiT) block for a joint spatial-channel attention mechanism into the MedSAM architecture. Within our cascaded framework, YOLOv11 first generates bounding boxes to localize the fetal head, serving as spatial prompts for MedSAM-AD to perform precise segmentation. The segmented fetal head is then processed with Canny edge detection and elliptical fitting to compute HC. Experimental results show that our approach achieves outstanding performance among standard biometric metrics of the HC18 dataset, attaining a Dice Similarity Coefficient (DSC) of 98.06 $\pm$ 1.06%, a Difference (DF) of 0.13 $\pm$ 2.47 mm, an Absolute Difference (AD) of 1.76 $\pm$ 1.74 mm, and a Hausdorff Distance (HD) of 1.18 $\pm$ 0.71 mm. With HD as the principal criterion for boundary delineation, our method achieves state-of-the-art performance in fetal head boundary delineation.

Keywords: Medical image segmentation · Fetal head circumference measurement · Deep learning

1 Introduction

Throughout pregnancy, ultrasound imaging plays a central role in fetal evaluation by detecting significant anomalies, monitoring growth trajectories, and assessing placental health, thereby supporting informed clinical decisions and interventions [7]. Measuring fetal head circumference (HC) is an essential anthropometric parameter in evaluating fetal health. By measuring fetal HC, doctors can predict gestational age and due date, and assess fetal development and mode of delivery [22]. The normal range of fetal HC reflects the brain development [16] and, therefore, plays an integral role in medical diagnosis.

D.-S. Huang et al. (Eds.): ICIC 2025, LNBI 15868, pp. 445–456, 2025.
https://doi.org/10.1007/978-981-95-0033-8_37

As shown in Fig. 1(a), existing methods for fetal HC measurement rely on an elliptical approximation of the head contour, where the contour's perimeter is calculated to estimate the HC value. Consequently, accurate contour detection is critical for HC estimation. Traditional manual measurement necessitates clinician expertise, exhibits time-consuming characteristics, and demonstrates precision constraints. Significant inter-observer variability exists among clinicians, compounded by a critical shortage of certified sonographers specialized in fetal ultrasound [5]. However, as illustrated in Fig. 1, fetal ultrasound images exhibit several inherent limitations. For instance, amniotic fluid and uterine wall structures mimic the texture and grayscale characteristics of the fetal head (see Fig. 1(b)); the fetal head boundaries are unclear (see Fig. 1(c)); and other elliptical-shaped tissues resemble the fetal head in morphology (see Fig. 1(d)). These artifacts result in incomplete contour detection or misidentification of non-cranial regions as the fetal head, posing significant challenges to HC measurement.

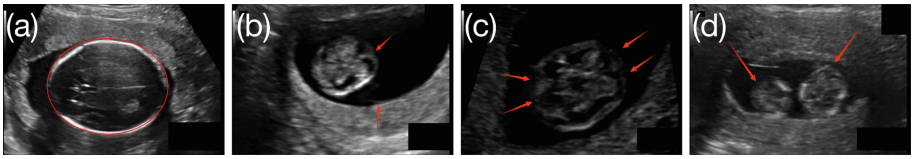


Fig. 1. Fetal ultrasound analysis: (a) Standard head measurement; (b) Boundary ambiguity from adjacent structures; (c) Incomplete edge continuity; (d) Tissue artifacts mimicking head contours.

Aiming to tackle these issues, we introduce YOSAM for automated fetal HC measurement in ultrasound images. YOLOv11 [11] is implemented for precise localization of the fetal head, thereby boosting segmentation performance. In addition, we integrate an Adapter layer [2] and a Dimensional Reciprocal Attention Mixing Transformer (D-RAMiT) [3] block into the image encoder of MedSAM [18], named MedSAM-AD. The use of the Adapter layer allows the adaptation of MedSAM for multimodality to the features of ultrasound images through lightweight parameters. At the same time, the D-RAMiT block captures local edges and global morphology in parallel through spatial and channel dual-branch attention. The synergy between the two significantly bridges the domain gap between MedSAM's pretraining data (CT/MRI) and fetal ultrasound-specific features. It effectively improves the segmentation ability of the MedSAM to ultrasound images.

The main contributions are summarized as follows:

- We use YOLOv11 to detect the fetal head and generate bounding box prompts, enhancing segmentation robustness against maternal tissue interference.
- We integrate an Adapter layer and a D-RAMiT block into MedSAM, which effectively improves the segmentation ability of ultrasound images.
- We apply YOSAM, a cascaded detection-segmentation framework, to the HC18 dataset and achieve a significant enhancement in HC measurement precision.

2 Related Work

In early research, methods such as Hough transform and machine learning have been employed to measure fetal HC. Lu et al. [17] presented a fully automatic fetal head detection and measurement method using the random forest classifier and Hough transform, enhancing robustness and accuracy in contour detection despite noise, artifacts, and poorly defined edges. A method combining random forest classifiers with fast ellipse fitting for HC measurement was proposed by Li et al. [13], enhancing accuracy and efficiency through the integration of prior knowledge and phase symmetry detection. However, these approaches face limitations in processing low-contrast images and complex anatomical structures.

In recent years, deep learning has driven rapid progress in marrying medical imaging with artificial intelligence. A notable development is the emergence of convolutional neural networks (CNNs), which enable automatic feature extraction from complex medical datasets, achieving exceptional performance in segmentation tasks. A regression CNN-based framework was proposed to directly estimate fetal HC from ultrasound images, achieving a mean absolute error of 4.52 mm without requiring manual segmentation or ellipse fitting [24]. This approach also highlights the potential for future improvements through attention mechanisms or multi-task learning. Li et al. [14] developed SAPNet, a dual-branch network that integrates segmentation and regression to simultaneously measure HC, biparietal diameter, and occipitofrontal diameter. Enhancing V-Net with attention mechanisms and deep supervision, DAG V-Net boosts the accuracy of delineating the fetal cranium and strengthens HC measurement stability [23]. Wang et al. [22] used GAC Net, a U-Net variant augmented with graph convolutions and SUO attention, achieving 98.21 $\pm$ 1.16% in Dice Similarity Coefficient (DSC) and 1.75 $\pm$ 1.71 mm in Absolute Difference (AD) on the HC18 dataset. In summary, advanced deep learning techniques have markedly elevated both the segmentation accuracy of the fetal head and the precision of HC estimation.

Benefiting from the remarkable success of Vision Transformer (ViT) [4] in image tasks, transformer-based architectures are increasingly adopted in medical image analysis. The Segment Anything Model (SAM) [12], built upon a ViT backbone, is a pioneering promptable framework that demonstrated exceptional zero-shot generalization capabilities across panoptic segmentation tasks through its advanced encoder-decoder design. Huo et al. [8] constructed DrSAM as an extension of SAM, retaining its pretrained weights while augmenting it with a U-shaped network and medical output tokens, offering a practical and adaptable tool for automated medical image analysis. MedSAM [18], an adaptation tailored for medical images, extends SAM's paradigm to achieve state-of-the-art performance in multi-organ segmentation. However, MedSAM performs suboptimally in fetal HC measurement tasks compared to specialized models. This limitation stems from its generalist design, which emphasizes multi-class segmentation robustness at the expense of domain-specific geometric precision. To address this limitation, we introduce YOSAM, a cascaded detection-segmentation framework integrating YOLOv11 and MedSAM-AD, optimized explicitly for fetal HC measurement in ultrasound images.

3 YOSAM

The automated workflow depicted in Fig. 2 outlines our approach for segmenting the fetal head and estimating HC. Firstly, the ultrasound image is passed through the image encoder to extract features, ultimately represented as image embeddings. At the same time, the fetal head is localized via a bounding box detection using the YOLOv11. Subsequently, the bounding box is encoded as positional prompts for the prompt encoder of MedSAM-AD, enabling targeted segmentation of the fetal head under positional guidance. Finally, since the segmentation result does not present a standardized ellipse shape, we need to perform edge detection on the segmentation result and fit the ellipse using least squares to approximate the standard ellipse contour. Based on the ultrasound imaging resolution, the HC metric is computed through pixel-to-millimeter conversion, with fetal head positioning derived from the centroid coordinates and major-axis orientation of the fitted ellipse.

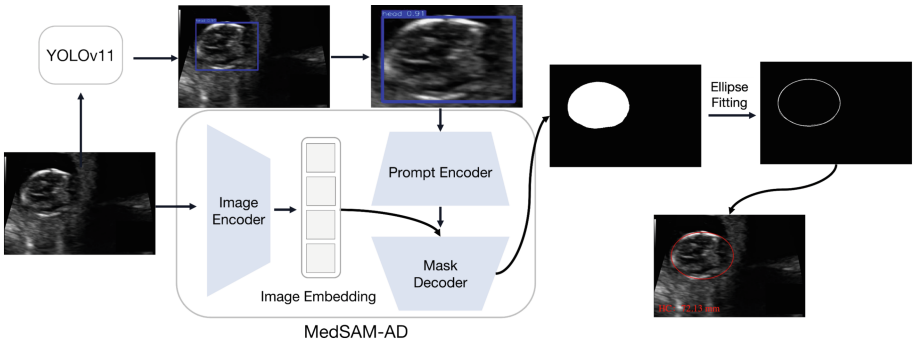


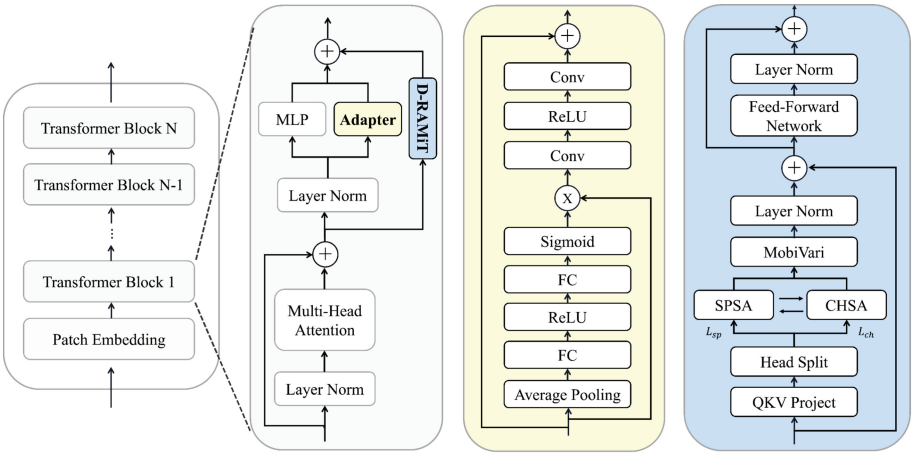
Fig. 2. Workflow of fetal head segmentation and HC measurement.

3.1 YOLO V11

YOLO (You Only Look Once) [11] is a groundbreaking real-time object detection framework that unifies region proposal and classification into a single neural network pass, enabling simultaneous prediction of bounding boxes and class probabilities during image processing. Building on this foundation, YOLOv11 introduces lightweight yet powerful architectural innovations to enhance efficiency and accuracy, enabling support for diverse tasks, including object detection, instance segmentation, and pose estimation [11]. In fetal HC measurement, YOLOv11 automates the detection of the fetal head and generates high-precision bounding box coordinates for MedSAM-AD, an interactive segmentation model requiring head localization data to produce segmentation masks. It can be found that bounding boxes deliver superior spatial contextualization for regions of interest, enhancing algorithmic precision in fetal head segmentation. By detecting fetal heads with high precision, YOLOv11 replaces manual annotation and feeds coordinates directly into MedSAM-AD's prompt encoder to enable consistent, efficient segmentation without human intervention.

3.2 MedSAM-AD

The MedSAM-AD architecture builds upon MedSAM, which has demonstrated robust performance in medical image segmentation [18]. As depicted in Fig. 2, the MedSAM-AD model comprises three primary modules: a prompt encoder, an image encoder, and a mask decoder. The role of the prompt encoder is to process user-specified bounding boxes into positional embeddings. The ViT-based image encoder extracts deep semantic features from medical images. Subsequently, the mask decoder receives both positional and image embeddings to generates segmentation results. Figure 3(a) illustrates how MedSAM-AD extends MedSAM by systematically integrating the Adapter layer and the D-RAMiT block within each transformer block of the image encoder.



(a) Image Encoder of MedSAM-AD (b) Adapter Layer (c) D-RAMiT Block

Fig. 3. (a) The architecture of the image encoder of MedSAM-AD. (b) Adapter layer architecture diagram. (c) D-RAMiT block architecture diagram.

Adapter Layer. According to [2], the integration of the Adapter layer into the SAM has been shown to significantly improve its segmentation accuracy in medical imaging tasks. The architecture of the Adapter layer is shown in Fig. 3(b). In MedSAM-AD, we integrate this adaptation mechanism into the image encoder to enhance ultrasound-specific feature representation. It implements feature adaptation for the image encoder along channel and spatial dimensions. For the channel dimension, the input feature map ($C \times H \times W$) is compressed to a size of $C \times 1 \times 1$ through global average pooling. These channel embeddings are passed through a linear layer to reduce their dimensionality to $\frac{C}{4} \times 1 \times 1$, then are reconstructed back to the original channel size by another linear layer. The channel weights are subsequently produced by a sigmoid activation layer. These weights undergo element-wise multiplication with the original input, and the product is propagated to the next network layer. For the spatial dimension, the input is spatially downsampled through a convolutional operation, after which the original resolution is recovered by transposing the convolution. To mitigate information loss, the Adapter

layer incorporates a residual connection by concatenating the initial input features with the processed output.

D-RAMiT Block. We incorporate the D-RAMiT block [3] (see Fig. 3(c)) into MedSAM to capture local edge details and global morphological features in ultrasound images. The D-RAMiT block processes input features through QKV projection, dynamically partitioned into two multi-head groups. Then, it operates through two parallel pathways: spatial self-attention (SPSA) and channel self-attention (CHSA), which jointly model local structural details and global channel-wise correlations. The parallel branches utilize distinct numbers of multi-head L_{sp} and L_{ch} to compute reciprocal attention weights. Specifically, SPSA leverages its attention heads to capture localized spatial relationships by incorporating embedded relative positional encodings, whereas CHSA employs the remaining heads to calculate cross-channel dependencies, thereby facilitating interactions between spatially distant yet semantically related patterns. A reciprocal modulation mechanism bridges consecutive blocks, whereby the value matrices of SPSA and CHSA undergo element-wise multiplication with spatially averaged features derived from the preceding CHSA and SPSA outputs, respectively. Following feature concatenation, the combined outputs are processed by MobiVari [3], which effectively mixes local and global attention. Output is generated by first applying LayerNorm, then feeding the result through a feed-forward network, and finally applying a second LayerNorm, each normalization step followed by a residual connection.

4 Experiments and Results

4.1 Dataset and Pre-Processing

HC18 dataset includes 1334 standardized plane ultrasound scans acquired from 551 normotrophic fetuses during routine prenatal examinations [6]. A total of 999 training images, each accompanied by manually delineated HC contours provided by specialists, are allocated to the training set, while 335 images are assigned to the testing set. This dataset encompasses ultrasound images from different trimesters of pregnancy. Specifically, the training set includes 165 images from the first trimester, 693 from the second trimester, and 141 from the third trimester. Correspondingly, the testing set comprises 55, 233, and 47 images from the first, second, and third trimesters, respectively. Each 2D ultrasound image measures 800 $\times$ 540 pixels, with spatial resolution varying between 0.052 and 0.326 mm per pixel.

To address the limited training data (n = 999) in the HC18 dataset and mitigate the risks of overfitting, we implemented a geometric augmentation protocol. Each ultrasound image underwent: 1) Random rotation simulates fetal head positional variance. 2) Horizontal flipping emulates left/right fetal presentation. 3) Vertical flipping accounting for probe orientation differences. This generated 20 augmented variants per original scan, expanding the training set to 19,980 images. Crucially, the testing set remained unaugmented to preserve evaluation integrity. The augmentation strategy alleviates data scarcity and enhances model robustness to anatomical presentation variability. To meet the model's input specifications, the images were resized to 1024 $\times$ 1024 pixels using bilinear interpolation.

4.2 Evaluation Metrics

Following the HC18 challenge evaluation criteria, segmentation performance was evaluated using the Dice Similarity Coefficient (DSC), HC Difference (DF), HC Absolute Difference (AD), and HC Hausdorff Distance (HD).

DSC quantifies the similarity between the segmentation result and the ground truth (GT), calculated as follows:

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \tag{1}$$

where A represents the segmentation result, and B denotes the GT. DSC is mathematically bounded within the closed interval [0,1], with values closer to 1 reflecting higher similarity between the segmentation and the GT.

DF reflects systematic measurement bias, and AD assesses the absolute error in HC measurements, calculated as follows:

$$DF = H_{pred} - H_{gt} \tag{2}$$

$$AD = |H_{pred} - H_{gt}| \tag{3}$$

where H_{pred} is the algorithm's predicted HC and H_{gt} is the true value. DF's positive and negative values indicate overestimation or underestimation trends, respectively. The values of AD directly reflect clinical usability, with lower values indicating greater measurement precision.

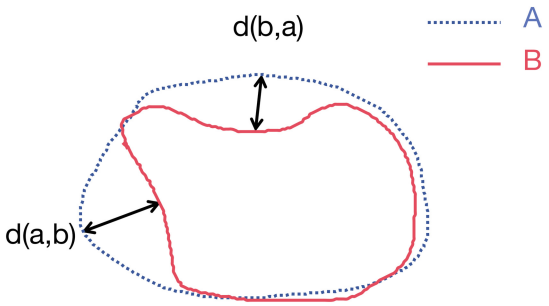


Fig. 4. Schematic illustration of Hausdorff Distance between point sets A and B.

HD quantifies the maximum spatial deviation between the GT and the segmented contour, calculated as follows:

$$HD(A, B) = max\left(\underset{a\in A}{maxmin}\, d(a, b), \underset{b\in B}{maxmin}\, d(b, a)\right) \tag{4}$$

where A represents the segmentation contour, and B denotes the GT. As illustrated in Fig. 4, $d(a, b)$ denotes the Euclidean distance between points a and b. A lower HD indicates a closer alignment of the segmented contour with the GT, reflecting minimal maximum local deviation.

4.3 Implementation Details

We initialized YOLOv11 with the official pre-trained parameters, subsequently fine-tuning the network on the HC18 dataset. MedSAM-AD was developed using PyTorch, with training on a single NVIDIA Quadro GV100 (32GB) GPU. The model adopted the official MedSAM pre-trained weights for initialization. During training, the training set was randomly divided into training and validation subsets with an 8:2 ratio. The Dice score was employed to assess segmentation performance on the validation set after each training epoch. For parameter optimization, the loss function consisted of an unweighted combination of dice loss and cross-entropy loss, with parameters updated via AdamW optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$). The learning rate was set to 1e-4 with a weight decay of 0.01. Due to GPU memory limitations, we set the batch size to 2 and trained the model for 20 epochs, with the entire training process taking 35 h.

4.4 Experimental Results

The segmentation accuracy was quantified on the HC18 testing set according to the criteria defined in Sect. 4.2, specifically: DSC, DF, AD, and HD. Our method achieved a DSC of $98.06 \pm 1.06\%$, a DF of 0.13 ± 2.47 mm, an AD of 1.76 ± 1.74 mm, and an HD of 1.18 ± 0.71 mm in the HC18 testing set. Our framework demonstrated distinct performance variations across trimesters. Notably, the model achieved high performance in fetal head segmentation during the second and third trimesters, with DSC values exceeding 98.15%, indicating excellent boundary delineation. Nevertheless, certain limitations persist; for example, the DSC was comparatively lower in the first trimester at 97.21%, and in the third trimester, the AD error reached 2.81 mm. Subsequent analysis revealed that the segmentation challenges encountered in the first trimester are primarily attributable to the dynamic nature of fetal head development. During this period, the skull and brain tissue remain insufficiently differentiated, and the ultrasound image features appear relatively blurred, diminishing the algorithm's sensitivity to boundary delineation. Conversely, the elevated AD error observed in the third trimester is associated with increased cranial calcification; the ultrasound images of the cranial region exhibit pronounced echo intensities and acoustic shadowing, which can impede the algorithm's capacity to accurately recognize actual boundaries.

Across all four evaluation metrics for fetal HC measurement, YOSAM outperforms established medical image segmentation baselines: U-Net [21], nnU-Net [9], Attention-UNet [15], V-Net [19], and MedSAM [18], with detailed comparative results documented in Table 1. Experimental results indicate that our framework outperforms these models across all four evaluation metrics in fetal HC measurement.

Table 1. Comparison results of YOSAM with different models.

Model	DSC(%)	DF(mm)	AD(mm)	HD(mm)
U-Net [21]	96.96 ± 5.74	1.65 ± 6.65	2.96 ± 6.18	1.71 ± 2.48
nnU-Net [9]	97.96 ± 1.13	1.29 ± 2.40	2.04 ± 1.80	1.21 ± 0.63
Attention-UNet [15]	97.91 ± 1.18	-1.05 ± 2.59	1.97 ± 1.98	1.28 ± 0.81
V-Net [19]	98.01 ± 1.06	1.16 ± 2.44	2.01 ± 1.80	1.21 ± 0.69
MedSAM [18]	97.91 ± 1.13	0.99 ± 2.55	2.05 ± 1.81	1.26 ± 0.70
YOSAM	**98.06 ± 1.06**	**0.13 ± 2.47**	**1.76 ± 1.74**	**1.18 ± 0.71**

To rigorously evaluate the performance of our framework, we carried out comparative analysis on the HC18 dataset against seven state-of-the-art fetal HC measurement algorithms from the literature [1, 6, 14, 16, 20, 22, 23]. As summarized in Table 2, our method ranks second in DSC, DF, and AD, closely trailing GCA-Net [22] in DSC/AD and DAG V-Net [23] in DF. Significantly, our framework establishes a new benchmark in terms of the HD metric with 1.18 ± 0.71 mm, surpassing all existing approaches. These results indicate that while GCA-Net marginally excels in region overlap accuracy and surface proximity, our method achieves clinically critical improvements in boundary delineation precision. It is worth noting that HD quantifies the greatest local discrepancy between the predicted boundary and the ground truth, making it crucial for detecting subtle yet critical edge inaccuracies that might be overlooked by overlap-based metrics like Intersection over Union (IoU) or DSC [10]. Even minor segmentation errors at the cranial boundary can lead to substantial HC miscalculations, potentially influencing clinical decisions. Our method is optimized to minimize extreme boundary errors, ensuring robust and clinically reliable segmentation outcomes.

Table 2. Comparison of different fetal HC measurement methods.

Model	DSC(%)	DF(mm)	AD(mm)	HD(mm)
Random forest [6]	97.10 ± 2.73	0.56 ± 4.21	2.83 ± 3.16	1.83 ± 1.60
GVF-Net [20]	95.53 ± 3.98	−0.24 ± 3.23	2.18 ± 2.40	2.42 ± 1.93
Mask-RCNN [1]	97.73 ± 1.32	1.49 ± 2.85	2.33 ± 2.21	1.39 ± 0.82
SAF-Net [16]	98.05 ± 4.02	1.26 ± 2.95	N/A	1.27 ± 0.77
SAPNet [14]	97.94 ± 1.34	0.59 ± 2.41	1.81 ± 1.69	1.22 ± 0.77
DAG V-Net [23]	97.93 ± 1.25	**0.09 ± 2.45**	1.77 ± 1.69	1.29 ± 0.79
GCA-Net [22]	**98.21 ± 1.16**	0.19 ± 2.32	**1.75 ± 1.71**	1.22 ± 0.71
YOSAM	98.06 ± 1.06	0.13 ± 2.47	1.76 ± 1.74	**1.18 ± 0.71**

4.5 Ablation Study

The ablation analysis of integrated Adapter layer and D-RAMiT block in the MedSAM framework is presented in Table 3, while comparative visual predictions are provided in Fig. 5. The baseline MedSAM attained a DSC of 97.91 ± 1.13%, with three spatial accuracy metrics: 0.99 ± 2.55 mm in DF, 2.05 ± 1.81 mm in AD, and 1.26 ± 0.70 mm in HD. When evaluated independently, the standalone D-RAMiT block improved boundary precision, achieving a DSC of 97.96 ± 1.07% while reducing DF to 0.86 ± 2.45 mm and HD to 1.22 ± 0.68 mm, demonstrating its effectiveness in mitigating extreme contour deviations. In contrast, integrating the Adapter layer alone slightly increased DSC to 97.99 ± 1.05% while significantly enhancing HC measurement robustness, as reflected in a reduced AD of 1.89 ± 1.72 mm. Most notably, the MedSAM-AD model, which incorporates the Adapter layer and D-RAMiT block, achieved the highest scores across all evaluation criteria. Specifically, it attained a DSC of 98.06 ± 1.06% and further reduced HC errors, with DF, AD, and HD values of 0.13 ± 2.47 mm, 1.76 ± 1.74 mm, and 1.18 ± 0.71 mm, respectively. Comparisons with configurations that utilize only a single module indicate that the combined approach yields superior improvements in all performance indicators. These results validate our hypothesis that cascaded detection-segmentation frameworks with domain-adaptive attention mechanisms are essential for overcoming ultrasound-specific artifacts in fetal biometry.

Table 3. Ablation study on the Adapter layer and D-RAMiT block in MedSAM.

Model	Adapter	D-RAMiT	DSC (%)	DF (mm)	AD (mm)	HD (mm)
MedSAM	✗	✗	97.91 ± 1.13	0.99 ± 2.55	2.05 ± 1.81	1.26 ± 0.70
	✗	✓	97.96 ± 1.07	0.86 ± 2.45	1.95 ± 1.71	1.22 ± 0.68
	✓	✗	97.99 ± 1.05	0.69 ± 2.46	1.89 ± 1.72	1.23 ± 0.71
	✓	✓	**98.06 ± 1.06**	**0.13 ± 2.47**	**1.76 ± 1.74**	**1.18 ± 0.71**

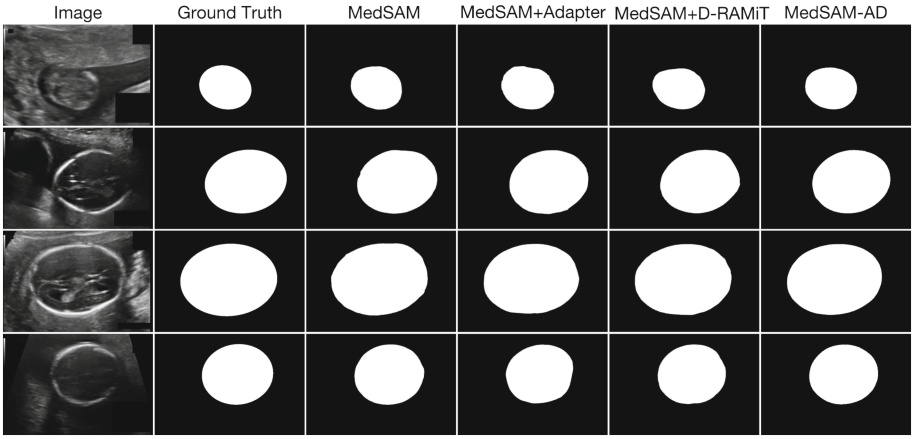


Fig. 5. Visualization of fetal head segmentation with different module integration.

5 Conclusion

We propose YOSAM, a cascaded framework that integrates object detection and segmentation for automated fetal HC measurement. Initially, YOLOv11 is employed to localize the fetal head region, after which MedSAM-AD performs precise segmentation. By incorporating an Adapter layer and a D-RAMiT block into the image encoder of MedSAM, the MedSAM-AD effectively adapts to the distinct features of ultrasound imaging, capturing both local edge details and global context, thereby leading to more precise fetal head segmentation. Comprehensive experimental evaluations reveal that our method delivers exceptional results across key metrics—including DSC, DF, AD, and HD—thereby establishing its potential as a robust solution for segmentation of the fetal head and as a reliable method for HC measurement. Future work will evaluate YOSAM on the other fetal ultrasound dataset, which includes multi-center and cross-device data, to further assess its generalizability.

Acknowledgments. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFF0606303.

References

1. Al-Bander, B., Alzahrani, T., Alzahrani, S., Williams, B.M., Zheng, Y.: Improving fetal head contour detection by object localisation with deep learning. In: Annual Conference on Medical Image Understanding and Analysis, pp. 142–150. Springer (2019)
2. Cheng, J., et al.: Sam-med2d (2023), https://arxiv.org/abs/2308.16184
3. Choi, H., et al.: Reciprocal attention mixing transformer for lightweight image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5992–6002 (2024)
4. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale (2021), https://arxiv.org/abs/2010.11929
5. Hekal, A.A., Amer, H.M., Moustafa, H.E.D., Elnakib, A.: Automatic measurement of head circumference in fetal ultrasound images using a squeeze atrous pooling unet. Biomed. Signal Process. Control **103**, 107434 (2025)

6. van den Heuvel, T.L., et al.: Automated measurement of fetal head circumference using 2d ultrasound images. PLoS ONE **13**(8), e0200412 (2018)
7. Horgan, R., Nehme, L., Abuhamad, A.: Artificial intelligence in obstetric ultrasound: a scoping review. Prenat. Diagn. **43**(9), 1176–1219 (2023)
8. Huo, X., Tian, S., Zhou, B., Yu, L., Li, A.: Dr-sam: U-shape structure segment anything model for generalizable medical image segmentation. In: International Conference on Intelligent Computing. pp. 197–207. Springer (2024)
9. Isensee, F., et al.: Nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021)
10. Karimi, D., Salcudean, S.E.: Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. IEEE Trans. Med. Imaging **39**(2), 499–513 (2019)
11. Khanam, R., Hussain, M.: Yolov11: an overview of the key architectural enhancements (2024), https://arxiv.org/abs/2410.17725
12. Kirillov, A., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026 (2023)
13. Li, J., et al.: Automatic fetal head circumference measurement in ultrasound using random forest and fast ellipse fitting. IEEE J. Biomed. Health Inform. **22**(1), 215–223 (2017)
14. Li, P., Zhao, H., Liu, P., Cao, F.: Automated measurement network for accurate segmentation and parameter modification in fetal head ultrasound images. Med. Biol. Eng. Comput. **58**, 2879–2892 (2020)
15. Lian, S., et al.: Attention guided u-net for accurate iris segmentation. J. Vis. Commun. Image Represent. **56**, 296–304 (2018)
16. Liu, P., Zhao, H., Li, P., Cao, F.: Automated classification and measurement of fetal ultrasound images with attention feature pyramid network. In: Second Target Recognition and Artificial Intelligence Summit Forum, vol. 11427, pp. 661–666. SPIE (2020)
17. Lu, W., Tan, J., Floyd, R.: Automated fetal head detection and measurement in ultrasound images by iterative randomized hough transform. Ultrasound Med. Biol. **31**(7), 929–936 (2005)
18. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nat. Commun. **15**(1), 654 (2024)
19. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
20. Rong, Y., et al.: Deriving external forces via convolutional neural networks for biomedical image segmentation. Biomed. Opt. Express **10**(8), 3800–3814 (2019)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015, Proceedings, Part III 18, pp. 234–241. Springer (2015)
22. Wang, X., Wang, W., Cai, X.: Automatic measurement of fetal head circumference using a novel gcn-assisted deep convolutional network. Comput. Biol. Med. **145**, 105515 (2022)
23. Zeng, Y., Tsui, P.H., Wu, W., Zhou, Z., Wu, S.: Fetal ultrasound image segmentation for automatic head circumference biometry using deeply supervised attention-gated v-net. J. Digit. Imaging **34**, 134–148 (2021)
24. Zhang, J., Petitjean, C., Lopez, P., Ainouz, S.: Direct estimation of fetal head circumference from ultrasound images based on regression cnn. In: Medical Imaging with Deep Learning, pp. 914–922. PMLR (2020)