Pubic Symphysis and Fetal Head Segmentation with LoRA Fine-Tuned SAM in Ultrasound Imaging

Marawan Elbatel[1], Robert Martí[2], and Xiaomeng Li[1]

[1] The Hong Kong University of Science and Technology
[2] Computer Vision and Robotics Institute, University of Girona

Abstract. This report presents the technical details of the team "Aloha" for Pubic Symphysis and Fetal Head Segmentation with Angle of Progression Estimation (PS-FH-AOP) MICCAI-2023 Challenge. Inspired by the recent Segment Anything Model (SAM) and the positional priors observed in transperineal ultrasound imaging of Pubic Symphysis and Fetal Head, we adapt SAM to ultrasound imaging. To preserve SAM's generalization capability while adapting to ultrasound imaging, we freeze the model and perform efficient low-ranked fine-tuning of the image encoder. The low-ranked fine-tuning can preserve the generalization capabilities of the pre-trained SAM model while being compact during inference. During inference, we integrate robust test-time augmentation techniques and employ a soft-ensemble approach based on 5-fold splits. The performance evaluation on the validation set demonstrates promising results, achieving a Dice coefficient of 92.9% and an 8.96 Hausdorff distance. Our code will be available at https://github.com/marwankefah/PS-FH-MICCAI-23

Keywords: SAM · LoRA · Medical Image Segmentation

1 Introduction

Fetal head descent and angle of progression (AOP) are considered reliable predictors for successful vaginal delivery. Estimated measurement usually relies on clinical examination with transperineal ultrasound. Time-consuming for real-time monitoring, the Angle of Progression (AOP) is determined by measuring the angle between two lines. The first line is drawn along the long axis of the pubic symphysis (PS), while the second line is drawn tangentially from the lower end of the pubic symphysis to the contour of the fetal head (FH). Thus, accurate segmentation of PS and FH is critical for the automatic estimation of the angle of progression

With the rise of large vision models, several works aim to intensify the search for models that exhibit strong performance while maintaining robust generalization capabilities in zero-shot [9], few-shot [10], and fully supervised scenarios [4,11]. Notably, prompt-driven segmentation models, such as the Segment

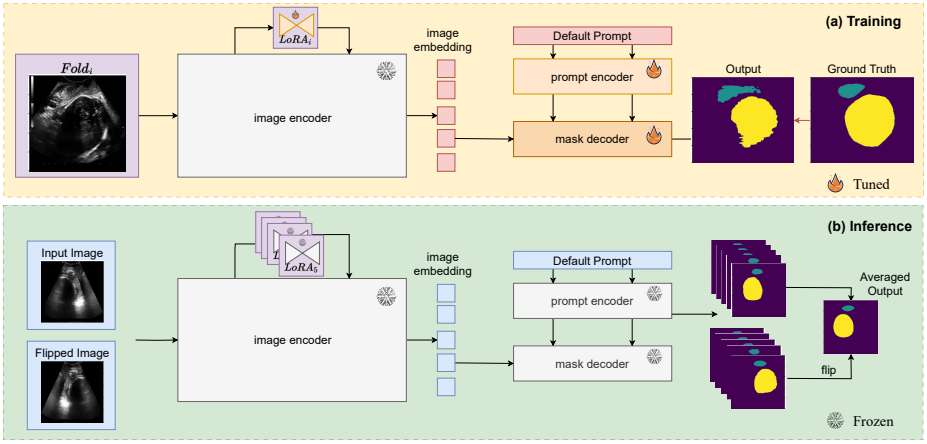

Fig. 1. The overall design of our training and inference pipeline

Anything Model (SAM) [5], have demonstrated significant advancements in numerous medical imaging tasks [8].

Nevertheless, the application of the Segment Anything Model (SAM) in ultrasound imaging poses a unique challenge. Ultrasound imaging is prone to high-frequency noise that is not encountered while training. This discrepancy between the noise characteristics of ultrasound imaging and the training data presents an obstacle in effectively leveraging SAM for ultrasound image segmentation.

To address this challenge, drawing inspiration from previous adaptations of SAM to medical imaging [11], we employ a low-ranked (LoRA) [3] fine-tuning strategy on SAM. Additionally, we enhance the robustness of our inference by utilizing a 5-fold soft-ensemble test-time augmentation approach. Fig. 1 (a) illustrates our proposed pipeline, which involves storing only the 5-fold LoRA parameters after training each fold in (a). Each set of parameters accounts for a mere 1% of the original model size (2GB) in terms of total model weights. During inference, we leverage these compact parameters in conjunction with soft ensembles and test-time augmentation techniques as depicted in Fig. 1 (b).

2 Methodology

The presence of essential positional priors and contextual relations between the pubic symphysis and the fetal head in transperineal ultrasound, as depicted in Fig. 2, highlights the appeal of efficiently fine-tuning general prompt-driven segmentation models. In this section, we outline the low-ranked fine-tuning strategy, LoRA [3], which we employ in our approach. We then describe the process of hyperparameter training. Finally, we provide an explanation of our inference pipeline.

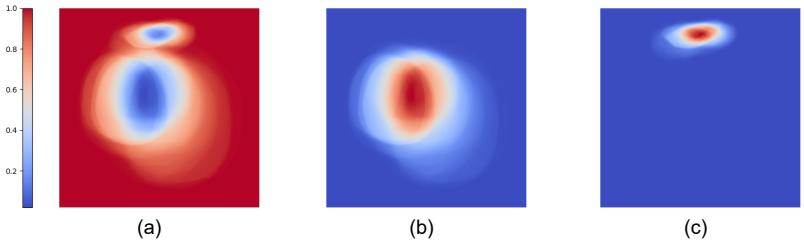

Fig. 2. The normalized heatmap of segmented object frequency in the original dataset provides information about the probability of different classes occurring at specific pixels in the training images. (a), (b), and (c) represent the background, fetal head, and pubic symphysis objects respectively.

2.1 Low-rank fine-tuning

For training on the dataset provided by the challenge organizer with a detailed version in [7]. We adopt the Low-rank fine-tuning strategy (LoRA) first introduced in [3].

To approximate the low-rank update of the parameters in the image encoder, we follow the approach proposed in [3]. Consider a model with a pre-trained weight matrix, denoted as W_0 with a forward pass output h as:

$$h = W_o X + \Delta W X \tag{1}$$

where ΔW represents the weight update, and X represents the input from the previous layer.

In this approximation, [3] propose to conduct updates through a low-rank matrix decomposition, given by

$$h = W_o X + \Delta W X = W_0 X + BAX \tag{2}$$

, where B is a matrix in $R^{d \times r}$, A is a matrix in $R^{r \times k}$, and the rank r is determined as a value less than the minimum dimension of d and k.

The final stored model parameters amount to 21M, which is just 1% of the size of the large vision model (2GB).

We repeat this process for 5 folds using the training set. It is worth noting that only the low-ranked parameters, preserving the original parameters of SAM, are saved. This ensures memory-efficient inference, making it straightforward to perform predictions with limited computational resources.

2.2 Training Recipe

For training SAM with LoRA, we utilize the vit-h model that has been pre-trained on natural imaging data, SA-1B [5], and freeze the parameters of the

entire image encoder. To ensure compatibility with the model, we up-sample the images to a resolution of 512x512.

Following the approach proposed in [11], we incorporate warm up technique [2] and employ the AdamW optimizer [6] during training. During each training epoch, a weighted sum of the cross-entropy loss and the dice loss is minimized. The dice loss is given a weight of 0.8. Each fold is trained for a total of 400 epochs.

To enhance the diversity of the training data, we apply various data augmentation techniques from the MONAI library [1], including horizontal flipping, Gaussian noise, blurring, random zooming, and affine transformation.

Finally, we set the rank of LoRA to be equal to four, as defaulted in [11].

2.3 Inference

During the inference stage, we utilize the vit-h model encoder with the frozen parameters. As a form of test time augmentation, we perform horizontal flipping, which effectively doubles the number of predictions obtained from each model. In total, we generate 10 predictions for each image by alternating between the 5-trained lightweight LoRA parameters.

By employing soft ensembles on the 10-generated predictions, we aim to reduce prediction variance and enhance robustness against perturbations in the input image space. Finally, we keep the largest connected components from each object as a postprocessing step.

References

1. Consortium, M.: Monai: Medical open network for ai (Jun 2023). https://doi.org/10.5281/zenodo.8018287, https://doi.org/10.5281/zenodo.8018287, If you use this software, please cite it using these metadata.
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
3. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021)
4. Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., Chen, R., Yu, J., Chen, J., Chen, C., Chi, H., Hu, X., Fan, D.P., Dong, F., Ni, D.: Segment anything model for medical images? ArXiv **abs/2304.14660** (2023), https://api.semanticscholar.org/CorpusID:258418033
5. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. arXiv:2304.02643 (2023)
6. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019), https://openreview.net/forum?id=Bkg6RiCqY7

7. Lu, Y., Zhou, M., Zhi, D., Zhou, M., Jiang, X., Qiu, R., Ou, Z., Wang, H., Qiu, D., Zhong, M., Lu, X., Chen, G., Bai, J.: The jnu-ifm dataset for segmenting pubic symphysis-fetal head. Data in Brief **41**, 107904 (2022). https://doi.org/https://doi.org/10.1016/j.dib.2022.107904, https://www.sciencedirect.com/science/article/pii/S2352340922001160
8. Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y.: Segment anything model for medical image analysis: An experimental study. Medical Image Analysis **89**, 102918 (2023). https://doi.org/https://doi.org/10.1016/j.media.2023.102918, https://www.sciencedirect.com/science/article/pii/S1361841523001780
9. Wald, T., Roy, S., Koehler, G., Disch, N., Rokuss, M.R., Holzschuh, J., Zimmerer, D., Maier-Hein, K.: SAM.MD: Zero-shot medical image segmentation capabilities of the segment anything model. In: Medical Imaging with Deep Learning, short paper track (2023), https://openreview.net/forum?id=iilLHaINUW
10. Wu, Q., Zhang, Y., Elbatel, M.: Self-prompting large vision models for few-shot medical image segmentation. ArXiv **abs/2308.07624** (2023), https://api.semanticscholar.org/CorpusID:260900153
11. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785 (2023)