Swin-DAG-VNet for Fetal Head Segmentation and Elliptical Parameter Regression for Circumference Measurement

Liyan Chen, Ye Ding(✉), Zhihao Li, and Li Lu

School of Computer Science and Technology, Dongguan University of Technology, Dongguan 523808, China
{241115300,dingye,241115318,221115220}@dgut.edu.cn

Abstract. Fetal head circumference (HC) is a key biometric indicator in prenatal ultrasound, essential for gestational age estimation and fetal growth assessment. However, conventional Convolutional Neural Networks (CNN)-based segmentation models often struggle to capture long-range dependencies, which hinders segmentation accuracy. To address this, we introduce Swin-DAG-VNet, a hybrid segmentation model which builds upon the Deeply Supervised Attention-Gated V-Net (DAG V-Net) as the baseline and integrates Swin Transformer to enhance global context modeling while preserving fine-grained structural details. Additionally, we incorporate Swin-Net-Add, a Transformer-enhanced feature fusion module, to improve multi-scale feature aggregation and boundary delineation. Furthermore, we employ an elliptical parameter regression method to predict key biometric parameters from the segmented contour, combines with an adaptive contour sampling strategy to refine segmentation, reducing noise and improving robustness. A physical calibration module ensures accurate real-world HC measurements. Experiments on the HC18 dataset demonstrate that Swin-DAG-VNet achieves an absolute difference (AD) of 1.78 mm, reducing absolute measurement bias by 5.3% compared to DAG V-Net, setting a new benchmark in the estimation of fetal HC.

Keywords: Fetal Head Circumference Measurement · Swin Transformer · Medical Image Segmentation · Elliptical Parameter Regression

1 Introduction

Fetal HC measurement plays a crucial role in prenatal ultrasound analysis, as it provides vital biometric information for gestational age estimation and fetal growth assessment. Accurate HC measurement is essential to detect abnormal growth patterns, such as intrauterine growth restriction (IUGR) and macrocephaly. Traditionally, HC measurements are performed manually by sonographers, introducing inter- and intra-operator variability, which can lead to measurement inconsistencies (Fig. 1). To address this challenge, automated fetal HC measurement methods have been developed to improve accuracy and reproducibility.

W. Senn et al. (Eds.): ICANN 2025, LNCS 16069, pp. 98–110, 2026.
https://doi.org/10.1007/978-3-032-04546-1_9

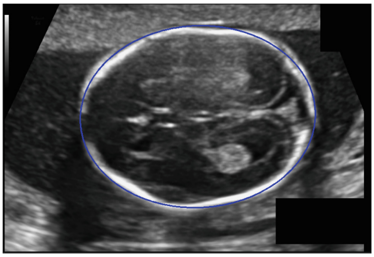

Fig. 1. Illustration of fetal HC measurement in a prenatal ultrasound image.

Recent deep learning-based segmentation models have significantly improved fetal head segmentation performance. Convolutional Neural Networks (CNNs), such as U-Net [12], Attention U-Net [7], and V-Net [10], have achieved promising results in medical image segmentation. However, these models primarily rely on local feature extraction with limited receptive fields, which restricts their ability to capture global contextual relationships. As a result, CNN-based models often struggle with fetal head segmentation challenges, including low contrast, shadow artifacts, and ambiguous boundaries in ultrasound images.

Recently, Vision Transformer (ViT) have demonstrated superior capability in modeling global context, outperforming CNNs across various medical imaging tasks [3]. Among them, the Swin Transformer [8], a hierarchical ViT variant, has exhibited notable advantages in capturing multi-scale contextual information while maintaining computational efficiency. However, directly applying the Swin Transformer to fetal ultrasound segmentation remains challenging due to the inherent characteristics of ultrasound imaging, including high noise levels, domain shifts, and substantial anatomical variations.

To address these limitations, we propose Swin-DAG-VNet, a novel hybrid segmentation model that integrates Swin Transformer and Swin-Net-Add [4] into DAG V-Net [18]. Swin-DAG-VNet leverages the strengths of both CNNs and Transformers: CNN-based local feature extraction ensures fine-grained anatomical preservation, while Transformer-based global attention improves segmentation robustness and structural consistency.

Furthermore, accurate HC measurement requires precise ellipse fitting, which is challenging due to segmentation noise and anatomical variability. Existing ellipse fitting approaches, such as Hough Transform [11] and Least Squares Method [5], often suffer from instability when applied to noisy ultrasound images. To address this limitation, we propose an elliptical parameter regression method, utilizing a multilayer perceptron (MLP)-based network to predict key ellipse parameters. Additionally, we introduce a dynamic contour sampling strategy, which adaptively selects high-confidence contour points, ensuring anatomically consistent HC estimation.

Our key contributions are summarized as follows:

- We propose ***Swin-DAG-VNet***, a novel Transformer-enhanced segmentation model that integrates Swin Transformer into DAG V-Net, significantly improving fetal HC segmentation accuracy.

- We introduce ***Swin-Net-Add***, a Transformer-based multi-scale feature aggregation module that enhances spatial consistency and boundary delineation.
- We propose an MLP-based ***elliptical parameter regression method*** and a ***dynamic contour sampling strategy*** to refine ellipse fitting, reducing segmentation-induced bias and ensuring anatomically consistent fetal HC measurements.

2 Related Work

Traditional medical image processing methods such as Edge Detection [15], Hough Transform [11], and Least Squares Method (LSM) [5] are initially used for fetal HC measurement. These methods detect object boundaries and fit geometric models to the fetal head contour. However, their reliance on handcrafted features makes them highly sensitive to noise, shadow artifacts, and ultrasound speckle patterns. Additionally, variations in fetal head shape and occlusions often lead to inaccurate measurements, limiting their clinical applicability.

CNN-based segmentation models have significantly advanced medical image analysis and improved the robustness of fetal head segmentation. U-Net [12] and V-Net [10] employ encoder-decoder architectures with skip connections to enhance feature representation. Variants like Pie-UNet [6] improve efficiency by reducing computational complexity but remain constrained by CNNs' inherent locality, limiting their ability to capture long-range dependencies in ultrasound images.

To mitigate this limitation, ***attention-based models*** such as Attention U-Net [7] and DAG V-Net [18] incorporate attention mechanisms to refine multi-scale feature representations. DAG V-Net enhances feature selection via attention gates [14], while ***deeper architectures*** like DEEPAM [20] employ hierarchical attention stacking to improve local feature extraction. However, these models still primarily focus on local context, which may hinder segmentation performance in challenging ultrasound images requiring global feature integration.

Recent advances in ***Transformer-based models*** address this issue by introducing self-attention for global context modeling. TransUNet [2] and Swin Transformer [8] extend Transformer-based architectures to medical image segmentation, while Swin-UNet [1] and MLFF-Transformer [19] improve boundary delineation via multi-scale feature interactions. However, despite their ability to model long-range dependencies, Transformer models often lack the strong local feature inductive bias necessary for precise boundary detection in medical images.

To bridge the gap between local and global feature extraction, we propose Swin-DAG-VNet, which integrates Swin Transformer into DAG V-Net. This hybrid approach enhances segmentation robustness and boundary delineation. Additionally, we introduce Swin-Net-Add, a residual fusion strategy for multi-scale feature integration, and a dynamic elliptical parameter regression method for precise HC measurement.

3 Methodology

3.1 Data Preprocessing

In the data preprocessing stage, we ***filled the labeled masks in the training set*** to enhance the model's ability to effectively learn the fetal HC region.

First, we loaded and converted the labeled mask images to grayscale to facilitate subsequent edge detection. A binarization process was then applied using a fixed threshold of 127, ensuring a clear segmentation of the target region and distinguishing the foreground (fetal head-enclosed region) from the background (non-enclosed region). Contour detection was subsequently employed to extract the edge information of the fetal HC from the binarized images. Finally, the resulting mask was filled with a uniform white value (255, 255, 255) to maintain consistency with the original annotation format during the training stage [17].

3.2 Swin-DAG-VNet

Figure 2 illustrates the workflow of Swin-DAG-VNet, our segmentation-based HC measurement model. The workflow begins with preprocessing, including image normalization and data augmentation, followed by segmentation using Swin-DAG-VNet, which extracts multi-scale features and captures long-range dependencies through the Swin Transformer-based encoder. The resulting segmentation mask is then used for elliptical parameter regression, ensuring anatomically consistent HC measurement.

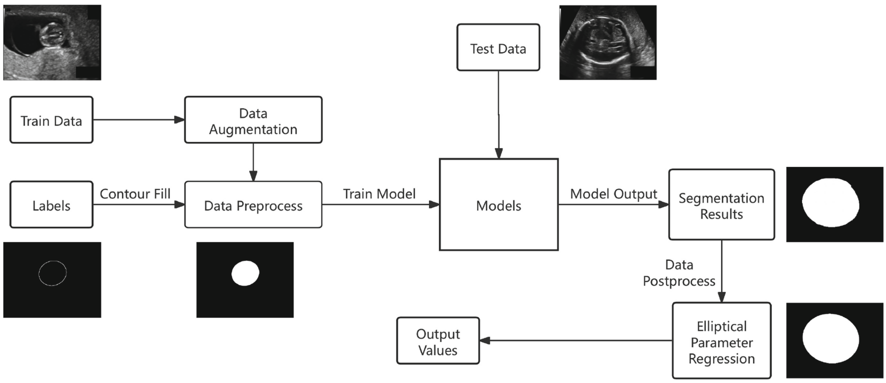

Fig. 2. Workflow of Swin-DAG-VNet for automated fetal HC measurement.

Unlike traditional CNN-based segmentation models, which rely on local convolutions, Swin-DAG-VNet introduces Transformer-based global feature modeling, effectively overcoming the limitations of CNNs in handling long-range dependencies and spatial inconsistencies. The architecture consists of three key components: ***Swin Transformer-based Encoder*** for global-local feature extraction. ***Swin-Net-Add Fusion Module*** for multi-scale feature aggregation.

Attention-Gated Decoder for precise contour delineation and HC segmentation. Swin-DAG-VNet follows an encoder-decoder architecture, similar to DAG V-Net, but replaces CNN-based residual concatenation with Swin Transformer-based fusion, as shown in Fig. 3. This modification enhances long-range dependency modeling and improves the accuracy of fetal head segmentation.

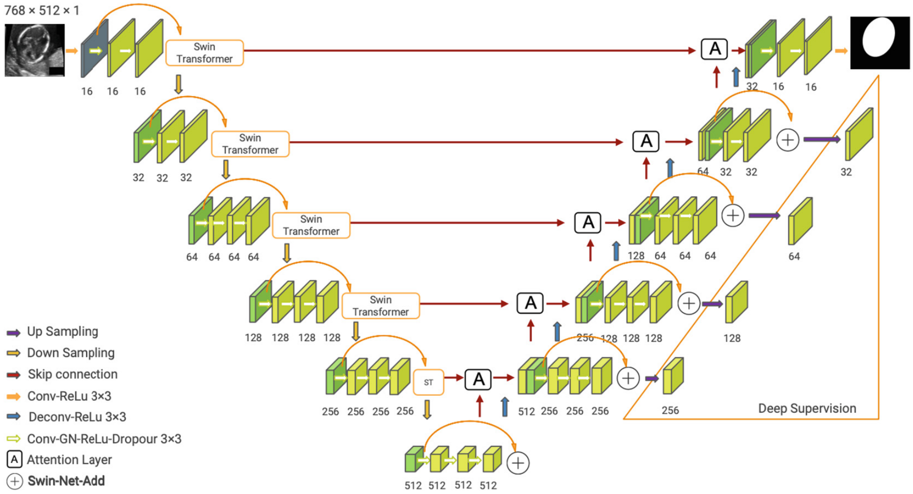

Fig. 3. Architecture of Swin-DAG-VNet for fetal head segmentation.

Swin Transformer-Based Encoder for Global Context Modeling. The Swin Transformer-based encoder replaces conventional CNN encoder to address long-range dependency modeling. Unlike CNNs, which have a limited receptive field, Swin Transformer leverages hierarchical self-attention mechanisms to efficiently capture both global and local contextual information [8].

The input ultrasound image is first partitioned into $P \times P$ non-overlapping patches, each mapped into a C-dimensional feature space via a linear projection:

$$X_{\text{patch}} = W_p \cdot X_{\text{input}} + b_p \tag{1}$$

where W_p and b_p are learnable projection parameters. This operation preserves local texture details while ensuring a structured feature representation.

Swin Transformer block comprises two components: Windowed Multi-Head Self-Attention, which performs self-attention within non-overlapping local windows to reduce computation, and the Shifted Window Mechanism, which enables cross-window interaction and mitigates information loss from window partitioning. Given an input feature map $X \in \mathbb{R}^{H \times W \times C}$, self-attention is computed as:

$$Y = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}} + B\right) V \tag{2}$$

where Q, K, V are the query, key, and value matrices derived from feature embeddings; d_k is the key dimension, ensuring numerical stability; B is the relative position bias, improving spatial awareness.

By combining hierarchical self-attention and cross-window communication, Swin Transformer-based encoding improves segmentation robustness.

Swin-Net-Add: Multi-scale Feature Aggregation. DAG V-Net originally employs ResNet-Add [16] for feature fusion. However, this method is limited to local feature interactions and cannot effectively model long-range dependencies in ultrasound images. To address this, we introduce Swin-Net-Add, a feature aggregation module that enhances multi-scale spatial consistency and improves fetal head segmentation accuracy [4].

Unlike the element-wise addition typically used in CNNs, Swin-Net-Add integrates Transformer-driven feature fusion, combining both global and local feature representations. This is achieved via: 1) Cross-window attention, which facilitates feature propagation across neighboring receptive fields, improving boundary delineation. 2) Feature fusion layers, which integrate high-level semantic information with low-level spatial details, ensuring morphological integrity.

To handle feature maps at different resolutions, a 1×1 convolution is first applied for channel alignment. The aligned feature maps are then processed through the Swin Transformer layer, where the self-attention mechanism computes long-range dependencies. Finally, feature aggregation is performed using:

$$F_{\mathrm{agg}} = W_1 \cdot F_{\mathrm{trans}} + W_2 \cdot F_{\mathrm{cnn}} \tag{3}$$

where F_{trans} represents Transformer-based global features; F_{cnn} represents CNN-based local features; W_1, W_2 are learnable weight parameters.

Attention Gated Decoder. The decoder employs an attentional gating (AG) mechanism [14] to suppress irrelevant regions during the upsampling process, ensuring a precise focus on the fetal head region. During decoding, deconvolution layers are used to progressively restore spatial resolution while incorporating skip connections from the encoder to enhance boundary recovery.

To further refine feature selection, an Attention Gating module is applied at each decoding stage, dynamically learning the weight distribution of salient feature regions to suppress background interference and improve segmentation accuracy. Finally, a Sigmoid activation function is utilized to generate the binarized segmentation mask, ensuring accurate delineation of the fetal head region. This framework effectively integrates spatial and contextual information, enhancing segmentation performance while mitigating extraneous influences.

3.3 Elliptical Parameter Regression Method

After segmenting the fetal head, the next crucial step is to accurately estimate the HC. Instead of traditional direct contour fitting methods, we propose an elliptical parameter regression method, which formulates HC estimation as a parametric regression problem [13].

Algorithm 1 details the complete procedure, including contour extraction, feature extraction, model training, and ellipse parameter prediction from the

Algorithm 1. Elliptical Parameter Regression

Require: Binary ultrasound image set $I \in \{0, 255\}^{H \times W}$ and corresponding ellipse parameters $P_{\text{true}} = (x, y, a, b, \theta)$
Ensure: Predicted ellipse parameters $P_{\text{pred}} = (x, y, a, b, \theta)$
Convert I to grayscale and apply binary thresholding: $I_{\text{bin}} = \text{Threshold}(I, 127)$
Extract contours $C = \text{FindContours}(I_{\text{bin}})$
Select the largest contour based on area: $C_{\max} = \arg\max(\text{Area}(C))$
for each image **do**
    Use the ground truth ellipse parameters P_{true}
    Perform adaptive sampling on $C_{\max}$ to obtain representative contour points:

$$C_{\text{sampled}} = \text{AdaptiveSampling}(C_{\max}, N)$$

    Construct feature vector $X \in \mathbb{R}^{2N}$ from the sampled coordinates
end for
Train a MLP with layers $512 \rightarrow 512 \rightarrow 256 \rightarrow 128$ and ReLU activations
Optimize using the loss function: ($\alpha = 0.6$ is a weighting parameter)

$$L = \alpha \sum_{i=1}^{N} ||P^i_{\text{true}} - P^i_{\text{pred}}||^2$$

for each test data X_{test} **do**
    Predict ellipse parameters:

$$P_{\text{pred}} = \text{MLP}(X_{\text{test}})$$

    Convert the predicted parameters to physical dimensions:

$$P_{\text{physical}} = (sx, sy, sa/2, sb/2, \theta)$$

    Adjust the orientation angle θ as follows:

$$\theta = \begin{cases} \theta + 90, & \text{if } \theta < 90 \\ \theta - 90, & \text{otherwise} \end{cases}$$

end for
Compute mean squared error (MSE) and intersection over union (IoU) to assess prediction accuracy
return Predicted ellipse parameters P_{pred}

segmented contours. An adaptive contour sampling strategy is employed to produce fixed-length, shape-representative feature vectors for ellipse regression by normalizing contours of varying lengths. For contours with ≥50 points (empirically chosen for performance), we uniformly sample 50 evenly spaced points, yielding sparser sampling for dense contours while preserving detail in simpler shapes. For shorter contours, points are duplicated to maintain consistent dimensionality. Although the method does not explicitly incorporate curvature or segmentation confidence, it effectively normalizes contour variability and provides a compact, informative representation for regression.

4 Experiments

Dataset Preparation and Augmentation. We use a dataset of 1334 two-dimensional (2D) ultrasound images from the Department of Obstetrics and Gynecology at Radboud University Medical Center, Netherlands. Each image is a resolution of 800×540 pixels, with pixel sizes ranging from 0.052 mm to 0.326 mm, depending on sonographer adjustments. To ensure a fair evaluation, the dataset is split into 75% for training and 25% for testing, ensuring intra-examination consistency [5].

To enhance training stability and prevent overfitting, ***extensive data augmentation and dropout were applied***. Each image was augmented 30 times with random horizontal or vertical flips and rotations between $-30°$ and $90°$, increasing the training set to 29,970 images. All images were resized to 768×512 to match the Swin-DAG-VNet input size. Dropout was also used during training for regularization. These strategies improved shape-invariant feature learning, ensuring robustness across varying fetal head orientations.

Implementation Details. To ensure stable training and fast convergence, network weights were initialized using Xavier initialization, with dynamic adjustments for activation functions (Sigmoid, ReLU, TanH) to maintain stable gradient propagation. The bias term was set to 0.1 to prevent large initial biases.

ReLU activation was applied to mitigate the vanishing gradient problem and accelerate convergence. The Adam optimizer was used with an initial learning rate of 5×10^{-4}, adjusted dynamically using an adaptive learning rate scheduler. The loss function combined Dice loss and cross-entropy loss to improve segmentation accuracy in the foreground region. The model was trained for 30 epochs with a batch size of 2, tuned based on GPU memory constraints.

5 Results

To comprehensively evaluate the effectiveness of Swin-DAG-VNet, we conducted a series of experiments under ***identical experimental conditions*** on the HC18 dataset.

To quantitatively assess segmentation performance, we employed HC absolute difference (AD, mm), Dice similarity coefficient (DSC, %), HC difference (DF, mm), and Hausdorff distance (HD, mm) as evaluation metrics.

Among these, AD and DF are the most clinically relevant metrics, as they directly measure the accuracy of HC estimation. AD quantifies the absolute difference between the predicted HC (HC_{pred}) and the ground-truth HC (HC_{gt}), AD is defined as:

$$AD = |HC_{\text{pred}} - HC_{\text{gt}}| \tag{4}$$

Meanwhile, DF evaluates the signed difference between HC predictions and ground truth, providing insight into systematic measurement biases:

$$DF = HC_{\text{pred}} - HC_{\text{gt}} \tag{5}$$

where positive DF values indicate overestimation of HC, and negative DF values indicate underestimation. A DF close to zero suggests minimal systematic bias, contributing to more reliable HC estimation.

Quantitative Comparison. Our experimental results demonstrate that Swin-DAG-VNet achieves the best AD and DF performance, significantly outperforming other models. Swin-DAG-VNet achieved an AD of 1.78 ± 1.71 mm and a DF of 0.11 ± 2.47 mm, indicating enhanced HC measurement accuracy while maintaining minimal systematic bias. Compared to DAG V-Net, Swin-DAG-VNet reduced AD by 5.3% and DF by 76.1%, ensuring more precise and unbiased HC estimation. Table 1 presents the results of the fetal HC segmentation and measurement on the HC18 test set for various deep learning models.

Meanwhile, to further assess the computational complexity and resource requirements of the proposed model, we also conducted a comparative analysis including the number of parameters, floating-point operations (FLOPs), and the inference time. The inference time listed in the table is the measurement based on a single-image inference, calculated by averaging 20 images from the HC18 test set. Swin-DAG-VNet integrates Transformer-based modules to enhance feature representation, which results in increased computational demands compared to convolution-based baseline models. However, despite not having the lowest parameter count or FLOPs, our model demonstrates a competitive inference time, highlighting its practical feasibility for real-world clinical deployment.

Table 1. Comparison of different segmentation models for fetal HC estimation.

Model	AD (mm)	DSC (%)	DF (mm)	HD (mm)	Params (million)	FLOPs (G)	Inference Time (second)
U-Net [12]	2.05 ± 1.92	95.62 ± 2.35	0.89 ± 2.67	2.38 ± 1.24	0.29	5.18	0.0352
nn-UNet [9]	1.98 ± 1.83	$\mathbf{98.02 \pm 1.05}$	1.22 ± 2.40	$\mathbf{1.18 \pm 0.65}$	19.07	412.65	11.65
Attention U-Net [7]	1.97 ± 1.98	97.91 ± 1.18	-1.05 ± 2.59	1.28 ± 0.81	31.38	32.40	14.26
V-Net [10]	2.01 ± 1.80	98.01 ± 1.06	1.16 ± 2.44	1.21 ± 0.69	45.60	676.23	6.04
DAG V-Net [18]	1.88 ± 1.68	97.99 ± 1.09	0.46 ± 2.49	1.22 ± 0.67	57.79	76.41	19.07
Swin-DAG-VNet	$\mathbf{1.78 \pm 1.71}$	97.92 ± 1.25	$\mathbf{0.11 \pm 2.47}$	1.27 ± 0.80	65.17	95.16	10.08

Figure 4 presents the results of the fetal head segmentation obtained using different models. Segmentation performance varied between models, with noticeable differences in contour smoothness and accuracy. Specifically, the U-Net model exhibited irregularities and jagged edges in the segmented fetal head contour, leading to a non-smooth shape. In contrast, Swin-DAG-VNet produced relatively smooth contours, with only minor irregularities in certain regions.

Ablation Study. To evaluate the contribution of each key component in Swin-DAG-VNet, we conducted a systematic ablation study by incrementally integrating Swin Transformer and Swin-Net-Add into the baseline DAG V-Net. The

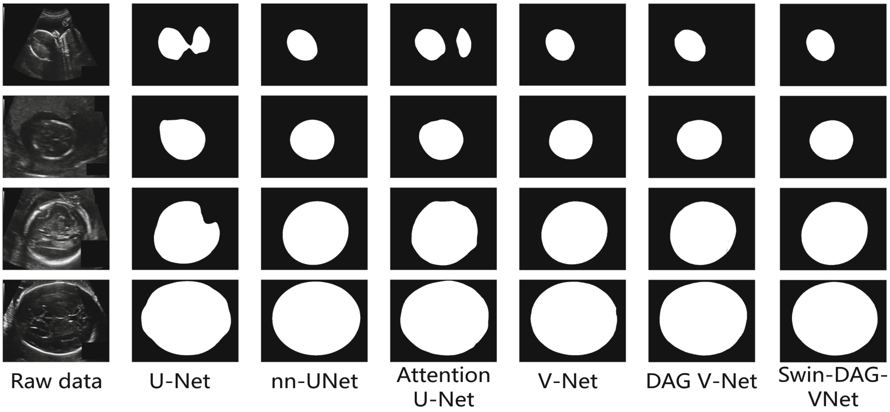

Fig. 4. Comparison of the segmentation results for fetal head.

results of the ablation study on the HC18 test set are presented in Table 2. The Swin Transformer module improved global feature extraction, reducing AD by 3.7% (from 1.88 mm to 1.81 mm) while enhancing segmentation stability, as indicated by a slight decrease in DSC from 97.99% to 97.93%. Additionally, the Swin-Net-Add module enhanced multi-scale feature aggregation and boundary delineation, leading to a substantial reduction in DF from 0.46 mm to -0.36 mm, thereby minimizing segmentation bias.

Table 2. Ablation Study on the impact of Swin Transformer and Swin-Net-add.

Model Variant	AD (mm)	DSC (%)	DF (mm)	HD (mm)
DAG V-Net (Baseline)	1.88 ± 1.68	97.99 ± 1.09	0.46 ± 2.49	$\mathbf{1.22 \pm 0.67}$
+ Swin Transformer [8]	1.81 ± 1.74	97.93 ± 1.24	0.13 ± 2.51	1.28 ± 0.83
+ Swin-Net-Add [4]	1.81 ± 1.75	$\mathbf{98.01 \pm 1.12}$	-0.36 ± 2.49	1.22 ± 0.72
Swin-DAG-VNet	$\mathbf{1.78 \pm 1.71}$	97.92 ± 1.25	$\mathbf{0.11 \pm 2.47}$	1.27 ± 0.80

Comparison of Ellipse Fitting Methods. Table 3 presents a comparative analysis of fetal HC measurement accuracy using different ellipse fitting methods, all evaluated under Swin-DAG-VNet. These results confirm that the elliptical parameter regression method achieves the lowest AD, improving HC measurement consistency and reducing segmentation-induced errors.

Table 3. Comparison of different ellipse fitting methods for HC measurement.

Method	AD (mm)	DSC (%)	DF (mm)	HD (mm)
Hough Transform [11]	1.96 ± 1.82	96.97 ± 2.17	$\mathbf{-0.06 \pm 2.67}$	1.90 ± 1.06
Least Squares Method [5]	1.81 ± 1.74	$\mathbf{98.00 \pm 1.12}$	-0.25 ± 2.50	$\mathbf{1.22 \pm 0.71}$
Elliptical Parameter Regression	1.78 ± 1.71	97.92 ± 1.25	0.11 ± 2.47	1.27 ± 0.80

Figure 5 shows a comparison of ellipse fitting results based on Swin-DAG-VNet. As shown, ellipses fitted by other methods exhibited angular shifts or image distortions in some cases. In contrast, the ellipses generated through elliptical parameter regression were smoother and more closely aligned with the segmentation results, demonstrating better consistency with the actual anatomical structure of the fetal head. These findings suggest that elliptical parameter regression enabled more accurate HC fitting, particularly in cases of irregular fetal head shapes, thereby reducing measurement bias caused by segmentation errors.

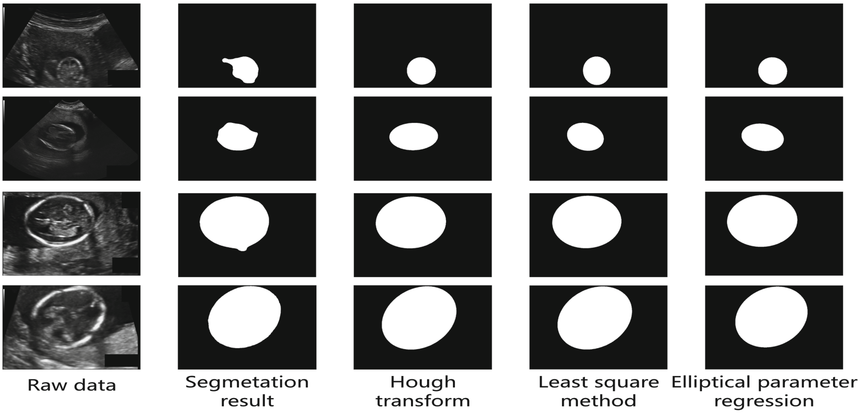

Fig. 5. Comparison of ellipse fitting results for Swin-DAG-VNet.

6 Conclusion

In this study, we propose Swin-DAG-VNet, a novel hybrid Transformer-enhanced model that integrates Swin Transformer and Swin-Net-Add module into DAG V-Net, effectively enhancing global context modeling while preserving local anatomical details. Swin-DAG-VNet surpasses DAG V-Net, achieving an absolute difference (AD) reduction of 5.3% (from 1.88 mm to 1.78 mm) and a measurement bias (DF) reduction of 76.1% (from 0.46 mm to 0.11 mm), demonstrating superior segmentation accuracy and robustness in fetal ultrasound images. Furthermore, we propose an elliptical parameter regression approach that employs an MLP-based network to predict key elliptical parameters, thereby reducing segmentation noise and ensuring anatomically consistent and accurate HC estimation. Future work will focus on enhancing segmentation performance in low-contrast fetal ultrasound images, exploring the potential of semi-supervised learning to improve model generalization, and validating Swin-DAG-VNet on larger, multi-center datasets to further assess its robustness and clinical applicability.

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

References

1. Cao, H., et al.: Swin-unet: unet-like pure transformer for medical image segmentation. In: Karlinsky, L., Michaeli, T., Nishino, K. (eds.) ECCV 2022. LNCS, vol. 13803, pp. 205–218. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-25066-8_9
2. Chen, J., et al.: Transunet: transformers make strong encoders for medical image segmentation (2021). https://arxiv.org/abs/2102.04306
3. Hatamizadeh, A., Yin, H., Heinrich, G., Kautz, J., Molchanov, P.: Global context vision transformers. In: International Conference on Machine Learning, pp. 12633–12646. PMLR (2023)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE, pp. 770–778 (2016)
5. van den Heuvel, T.L., de Bruijn, D., de Korte, C.L., van Ginneken, B.: Automated measurement of fetal head circumference using 2D ultrasound images. PloS One **13**(8), e0200412 (2018)
6. Jiang, Y., Zhang, X., Chen, Y., Yang, S., Sun, F.: Pie-UNet: a novel parallel interaction encoder for medical image segmentation. In: Iliadis, L., Papaleonidas, A., Angelov, P., Jayne, C. (eds.) ICANN 2023. LNCS, vol. 14255, pp. 558–569. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-44210-0_45
7. Lian, S., Luo, Z., Zhong, Z., Lin, X., Su, S., Li, S.: Attention guided u-net for accurate iris segmentation. J. Vis. Commun. Image Represent. **56**, 296–304 (2018)
8. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. In: ICCV, pp. 10012–10022. IEEE/CVF (2021)
9. Luu, H.M., Park, S.H.: Extending nn-unet for brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) BrainLes 2021. LNCS, vol. 12963, pp. 173–186. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-09002-8_16
10. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 3DV, pp. 565–571. IEEE (2016)
11. Mukhopadhyay, P., Chaudhuri, B.B.: A survey of Hough transform. Pattern Recogn. **48**(3), 993–1010 (2015)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Sarstedt, M., Mooi, E., Sarstedt, M., Mooi, E.: Regression analysis. A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics, pp. 209–256 (2019)
14. Schlemper, J., et al.: Attention gated networks: learning to leverage salient regions in medical images. Med. Image Anal. **53**, 197–207 (2019)
15. Sun, R., et al.: Survey of image edge detection. Front. Signal Process. **2**, 826967 (2022)
16. Xu, W., Fu, Y.L., Zhu, D.: Resnet and its application to medical image processing: research progress and challenges. Comput. Methods Programs Biomed. **240**, 107660 (2023)
17. Xu, Z., Baojie, X., Guoxin, W.: Canny edge detection based on open CV. In: IEEE/ICEMI, pp. 53–56. IEEE (2017)
18. 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)

19. Zhang, J., Fang, Z., Wang, Z.: Multi-feature fusion enhanced transformer with multi-layer fused decoding for image captioning. Appl. Intell. **53**(11), 13398–13414 (2023)
20. Zhong, S., Wen, W., Qin, J., Huang, Z.: DEEPAM: toward deeper attention module in residual convolutional neural networks. In: Wand, M., Malinovská, K., Schmidhuber, J., Tetko, I.V. (eds.) ICANN 2024. LNCS, vol. 15016, pp. 405–418. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72332-2_26