Received: 29 October 2023 | Revised: 11 December 2023 | Accepted: 14 January 2024

DOI: 10.1002/ima.23031

WILEY

RESEARCH ARTICLE

Automatic measurement of fetal abdomen subcutaneous soft tissue thickness from ultrasound image based on a U-shaped attention network with morphological method

Zhenming Yuan[1] | **Tianhao Xu**[1] | **Cheng Yu**[2] | **Xiaojun Ye**[2] | **Jian Zhang**[1]

[1]School of Information Science and Technology, Hangzhou Normal University, Hangzhou, China

[2]Hangzhou Women's Hospital, Hangzhou, China

Correspondence
Jian Zhang, School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China.
Email: jeyzhang@hznu.edu.cn

Funding information
Primary Research and Development Plan of Zhejiang Province, Grant/Award Number: 2020C03107; National Natural Science Foundation of China, Grant/Award Number: 61972361

Abstract

Fetal abdominal subcutaneous soft tissue thickness (FASSTT) is a key indicator in evaluating fetal growth, development, and nutritional status. Currently, manual measurement in FASSTT faces the problems of difficulty in positioning, time consumption, and inaccurate measurement. Therefore, this article proposes an automatic measurement scheme for FASSTT. Firstly, a U-shaped attention network VGG-SeUnet is proposed to automatically segment the subcutaneous soft tissue area of the fetal abdomen. Secondly, based on the segmentation results, a morphological method is proposed to obtain FASSTT. Specifically, the segmentation network uses VGG16 as the encoder and connects the decoder through jump connections for multi-scale fusion. We introduce channel attention to jump connections, which enables the model to select key channels for feature fusion. At the same time, the model uses the proposed DF_Loss function for training to solve the problem of sample imbalance. Based on the segmentation results, a morphological distance transformation algorithm is proposed to obtain FASSTT by drawing the maximum inscribed circle and calculating its diameter. The method is evaluated on a fetal abdominal circumference ultrasound dataset containing 135 samples. The DSC, mIOU, mPA, Precision and Recall of the segmentation experiments reache 88.9%, 81.02%, 90.44%, 86.38%, and 90.44% respectively. The difference between FASSTT's result from that provided by radiologist is 0.0615 cm in Average Error (AVE) and 0.081 cm in Root Mean Square Error (RMSE). The overall performance of this algorithm is superior to existing methods with excellent performance, and the measurement error is less than 1 millimeter. This result proves that the proposed scheme in this paper can achieve automated measurement of FASSTT and assist doctors in subsequent evaluation of fetal development and nutritional status.

KEYWORDS

automated measurements, convolutional neural networks, fetal abdomen subcutaneous soft tissue thickness, segmentation, ultrasound imaging

1 | INTRODUCTION

Fetal growth and development in the uterus is a complex process, and imbalanced intrauterine nutrition (including intrauterine malnutrition and overnutrition) can lead to fetal underdevelopment and have a negative impact on fetal growth.[1,2] However, traditional examination methods are difficult to directly obtain the growth and development status of the fetus in the uterus. Therefore, ultrasound imaging has become an important clinical method for evaluating fetal growth because it is non-invasive, non-radioactive, and easy to perform.[3] In previous studies, the fetal abdomen subcutaneous soft tissue thickness (FASSTT) has been found to be significantly correlated with the actual birth weight (ABW) of the newborn.[4–6] Therefore, measurement of FASSTT in ultrasound image is significant for evaluating fetal nutritional status.

The current method of manually measuring FASSTT firstly locates the subcutaneous soft tissue area of the fetal abdomen to be measured, and then measures the length of the thickest part of the area, which is FASSTT. However, the subcutaneous soft tissue area of the fetal abdomen only accounts for 10% of the entire abdominal circumference (AC) ultrasound image, with a thickness of only 4–7 mm. In ultrasound images with a resolution of 96dpi, manual measurement annotation errors of 3–4 pixels will result in an error of 1 mm, which reaches 14.3%–25% of the entire thickness. The current manual measurement methods largely depend on the experience of doctors. Therefore, achieving rapid automatic localization and accurate measurement of fetal subcutaneous soft tissue thickness is of great clinical significance.

In recent years, deep learning methods, especially convolutional neural networks (CNN), have been successfully applied to fetal biometrics. The current research mainly focuses on automatic measurement of fetal head circumference (HC), biparietal diameter (BPD), femoral length (FL), and AC,[7] and it still lacks effective automatic solutions for FASSTT measurement. The reason is two-fold. First, due to the complexity of fetal organs and tissues, the subcutaneous soft tissue region could not be observed with clear boundaries and there was a sample imbalance problem that means the pixels in the tissue region (positive samples) were massively smaller than those in the background region (negative samples). This problem makes segmentation and localization of this region difficult. Secondly, there is a lack of effective measurements to find the thickest part of the subcutaneous soft tissue region. To this end, we propose a U-shaped attention network VGG-SeUnet with morphological method for FASSTT measurement. First, the subcutaneous soft tissue region of the fetal abdomen is automatically segmented by the VGG-SeUnet network. The encoder part of the network performs feature extraction via VGG-16 and connects to the decoder via jump connections with channel attention (Se-Blocks[8]) to form a multilevel U-shaped structure. Secondly, a morphology-based distance transformation algorithm is proposed to find the minimum distance (MD) from each non-zero pixel inside the segmented tissue to the zero pixel outside of the tissue. The maximum value of MD is the radius of the maximum inscribed circle inside the tissue, and we take the diameter as the FASSTT.

It is worthwhile to highlight several aspects of the proposed method:

1. This paper proposes a new U-shaped Attention Network-based fetal abdominal subcutaneous soft tissue region segmentation network named VGG-SeUnet. We use VGG16 as the encoder and the Unet upsampling structure as the decoder, and connect them by a jump connection for better feature extraction capabilities. A channel attention mechanism (Senet) is added to the jump connection to enhance the fusion of shallow and deep features, reduce the number of parameters, and model the complex correlation between channels. A 1×1 convolution operation is performed in the classifier to reduce the channel and enable cross-channel information interaction.
2. The loss function DF_LOSS is constructed to solve the problem of sample imbalance by combining similarity calculation and giving different weights to difficult and easy samples. This overcomes the shortcomings of the similarity-based loss function and improves the overall segmentation performance of the model.
3. The distance transformation algorithm can not only measure the thickness of the fetal abdomen subcutaneous soft tissue but also indicate the location of the maximum inscribed circle within soft tissue area. By plotting the maximum inscribed circle, we provide intuitive auxiliary information for the final decision of doctors.

The rest of the paper is organized as follows: Section 2 summarizes the related works. Section 3 introduces the VGG-SeUnet, DF_LOSS loss function, and distance transformation-based maximum inscribed circle location in detail. Section 4 demonstrates the experimental results and discussion is shown in Section 5. Section 6 concludes the paper.

2 | RELATED WORK

In the study of automatic measurement of fetal bioindicators, several methods for automatic measurement of fetal HC and BPD in ultrasound images have been proposed in recent years. Zeng et al.[9] proposed a deeply supervised attention-gated (DAG) V-Net network that enhances the V-Net model with attention and trains the model through multi-scale loss function for automatic segmentation of the fetal head from ultrasound images, which is followed by morphological processing with ellipse fitting to perform HC measurements. Zeng et al.[10] proposed to incorporate separable convolutions into a sequential prediction network to improve the speed and accuracy of the fetal head segmentation. Morphological processing with least squares ellipse fitting is subsequently conducted to obtain fetal HC. Oghli et al.[11] proposed a deep learning-based fetal head segmentation method for automatic BPD and HC measurement, which overcomes the weakness of U-net by multi-feature pyramidal scheme (MFP). This model was validated on the HC18 challenge dataset and demonstrated superior performance on automatic measurement of BPD and HC. Similarly, ultrasound image-based biometry has been studied in femur length and AC. Zhu et al.[12] proposed a deep learning approach to automatically measure the fetal femur length. They used SegNet network to segment the fetal femur and obtained the femur length by calculating the intersection of the skeletal curve of the segmented femur with the femur contour and locating the final endpoint of the femur. Compared with random forest regression model, SegNet yields much better performance. Kim et al.[13] used a hierarchical identification method to correlate multiple CNN models to perform AC measurements after semantic segmentation from 2D ultrasound image. The validity of the proposed method was verified by clinical data. A Dice similarity measure of 92.55 ± 0.83 was achieved in AC measurements, with an accuracy of 87.10% in the acceptance examination of the standard planes of the fetal abdomen.

While the above researches have shown good results in automating the measurement of the biological parameters of fetal BPD, HC, AC, and FL, there is still a lack of effective automated measurement method for the measurement of subcutaneous soft tissue thickness in the fetal abdomen. The segmentation and measurement of abdominal subcutaneous soft tissue thickness still face great challenges due to the complex surrounding tissues, unclear edges, uneven sample distribution, and lack of effective morphological measurement methods. Therefore, it is of great research value and clinical significance to propose a method that can accurately measure subcutaneous soft tissue thickness in the fetus.

3 | METHOD

Inspired by the clinical manual measurement method of FASSTT, this paper proposes an automatic FASSTT measurement method consisting of two modules. The first module uses a new VGG-SeUnet model to extract multi-scale features from ultrasound images. Training is performed through the proposed DF_Loss function to achieve automatic segmentation of subcutaneous soft tissue areas and solve the problem of sample imbalance. VGG-SeUnet utilizes VGG16[14] as the encoder for feature extraction and gradually fuses its intermediate output with the decoder to upsample the features to the original size. This is a typical U-net[15] structure. In order to enhance the interaction between different scales, this paper introduces a channel attention module (SeNet) in the feature fusion skip connection and introduces dimensionality reduction convolution in the classifier. The second module applies morphological methods to the segmented ROI area to find the location with the maximum thickness of the subcutaneous soft tissue area of the fetal abdomen, and draws the maximum inscribed circle to visually display the measurement location, and finally obtains the FASSTT. The overall framework is shown in Figure 1.

3.1 | VGG-SeUnet

It is very difficult to automatically segment fetal ablevel-dominal subcutaneous soft tissue from ultrasound images because different organs and tissues are intricately intertwined and it is difficult to determine tissue boundaries. At the same time, highly complex networks may not be appropriate for this task because the lack of rich training samples may lead to overfitting. To this end, we propose a U-shaped attention network VGG-SeUnet with skip connections to extract multi-scale feature information reflecting high-level and low-level semantics of tissue segmentation. In addition, the channel attention in skip connections and the dimension reduction in classifier enhances multi-level feature fusion and reduces the amount of parameters to lower the risk of overfitting. The detailed structure of VGG-SeUnet is shown in Figure 2.

In the feature extraction network, this paper divides VGG16 into 5 modules (Block1-5) as the encoder part. The decoder part also contains 5 Blocks with one-to-one correspondence with the encoder blocks. The encoder block at each layer is connected to the decoder counterpart through jump connection. We use pre-trained weights to initialize the multi-layer encoder to enhance the multi-level feature learning ability. The VGG16

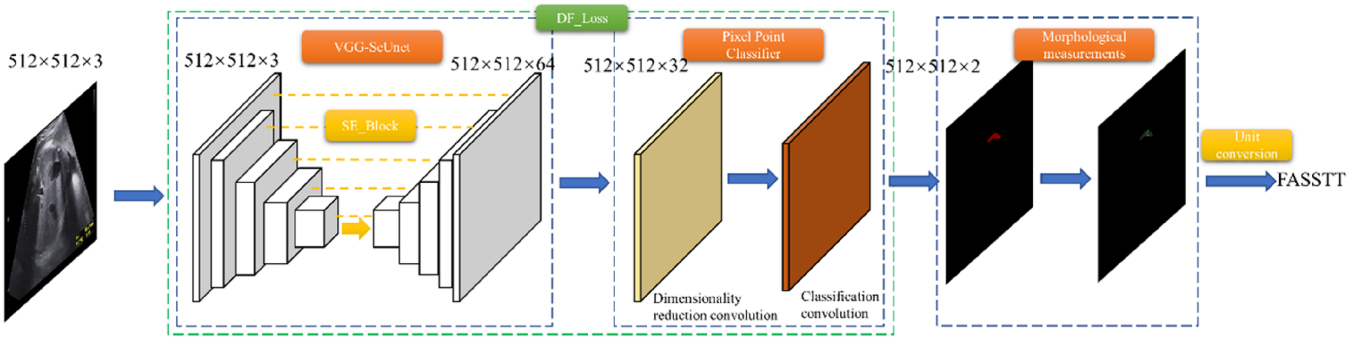

FIGURE 1 The framework of the proposed approach.

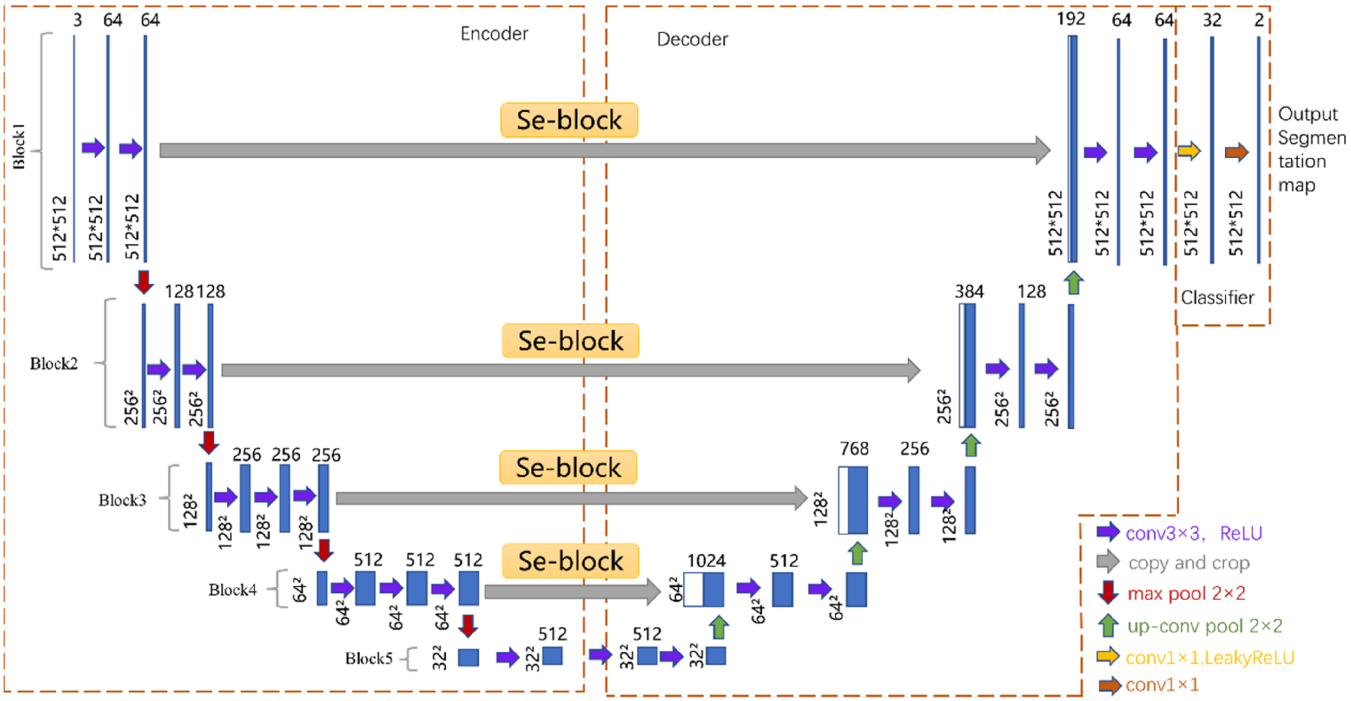

FIGURE 2 VGG-SeUnet network structure in which purple arrow indicates 3 × 3 convolution with ReLU activation function, gray arrow indicates jump connections, red arrow indicates down-sampling, green arrow indicates up-sampling, yellow arrow indicates 1 × 1 convolution with Leaky ReLu activation function, and orange arrow indicates 1 × 1 convolution.

network consists of convolutional layers and pooling layers. The network input data is the image of (512, 512, 3). In Block1, the input image is processed twice with 3 × 3 convolution and ReLu activation function to obtain the feature maps of (512, 512, 64), which are down-sampled through a 2 × 2 max pooling layer to yield the input of Block2 with size (256, 256, 128). Block2 is the same as Block1. The input of Block3 obtained by downsampling is of size (128, 128, 256). Each of Block3, Block4 and Block5 performs three 3 × 3 convolution and ReLu activation based on the input data, and the output feature maps of Block5 with size (32, 32, 512) are fed to the decoder part. In each layer of the decoder, the feature maps are upsampled through transposed convolution and passed to the upper layer where the features are merged with the features obtained by the encoder block at this layer through skip connection.

Considering that the convolution operation[16] inevitably introduces information decay and different channels may contribute differently to the segmentation task, we therefore introduce a channel attention scheme for re-weighting the new feature maps during each jump connection so as to highlight the channels with key information and suppress those without key information for better feature fusion. Actually, the channel attention is realized through Se-Block whose structure is shown in Figure 3. Firstly, each feature map is compressed by global average pooling to obtain a 1 × 1 × C vector, which is normalized

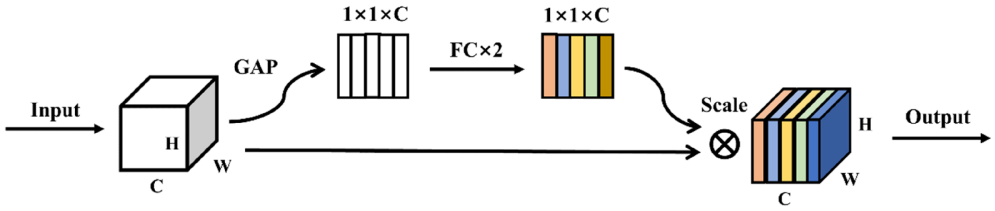

FIGURE 3 Se-Block in which GAP denotes the aggregation method through global average pooling (GAP), FC × 2 denotes two fully connected layers to produce the channel weights, and Scale denotes the point-wise multiplication of the weights to the input channels to achieve recalibration of the channels.

by two fully connected layers with sigmoid activation to obtain the weight of each channel. Subsequently, the recalibration of the channels is achieved by multiplying the weights with the original input channels.

In order to reduce the computational complexity of the network and reduce the feature dimension, we perform 1 × 1 convolution on the output classifier part to reduce the dimension from 64 channels to 32, and use the LeakyReLU function for activation before outputting the required number of categories. This allows the network to further reduce the complexity of operations while maintaining high-level features, and improve the segmentation performance and generalization ability of the model.

3.2 | DF_Loss function

Automatic measurement of FASSTT relies on accurate distinction between the abdominal subcutaneous soft tissue area and the background area. However, the number of pixels in the tissue region to be segmented (positive samples) is significantly less than the number of pixels in the background region (negative samples). The problem of sample imbalance will decrease the applicability of traditional cross-entropy loss function to the training of segmentation networks. Dice loss function[17] considers only positive samples by evaluating the ratio between the intersection of the predicted with true tissue region and the union of the two regions such that the sample imbalance problem is avoided. However, dice loss does not consider the extent to which the predicted tissue pixel deviates from the ground truth. Focal loss function[18] mitigates the problem by assigning different weights to the pixels that are highly probable to be misclassified and pixels that tend to be correctly classified, thus emphasizes the performance improvement from the pixels that are hard to distinguish. In addition, focal loss also adopts weighting mechanism to deal with large discrepancy in the number of positive and negative samples.

Based on the aforementioned reasons, we combine the dice loss function and the focus loss function to construct a DF_Loss to guide the training of VGG-SeUnet network. Specifically, let $X^p = \{x_i^p\}_{i=1,\ldots,N_p}$ be the predicted tissue region where x_i^p denotes pixel, $Y^g = \{y_i^g\}_{i=1,\ldots,N_g}$ be the ground truth tissue region defined by annotation and $X = \{x_i\}_{i=1,\ldots,N}$ be all the pixels in the ultrasound image. The dice loss function can be denoted as follows:

$$L_{dice} = 1 - \frac{2\,|X^p \bigcap Y^g\,|}{|X^p| + |Y^g|}, \tag{1}$$

where $|\cdot|$ computes the cardinality of a set. The focal loss function can be represented as follows:where $c\epsilon\{1,0\}$ is the ground truth label, $P(x_i)$ is the probability that x_i belongs to the tissue, γ is a hyper parameter controlling the contribution of the error-prone pixels to the classification loss value and α is also a hyper parameter controlling the contribution of positive and negative samples that can be computed through $\frac{\alpha}{1-\alpha} = \frac{N-N_g}{N_g}$. It is obvious that if Y_g accounts for very small region of the image, $\frac{N-N_g}{N_g}$ will be very large such that α is also large. This guarantees that the pixels potentially correlated to the

$$L_{focal} = \frac{\sum_{i=1}^{N}(-\alpha \cdot c \cdot (1-P(x_i))^{\gamma}\log(P(x_i)) - (1-\alpha)(1-c)P(x_i)^{\gamma}\log(1-P(x_i)))}{N}, \tag{2}$$

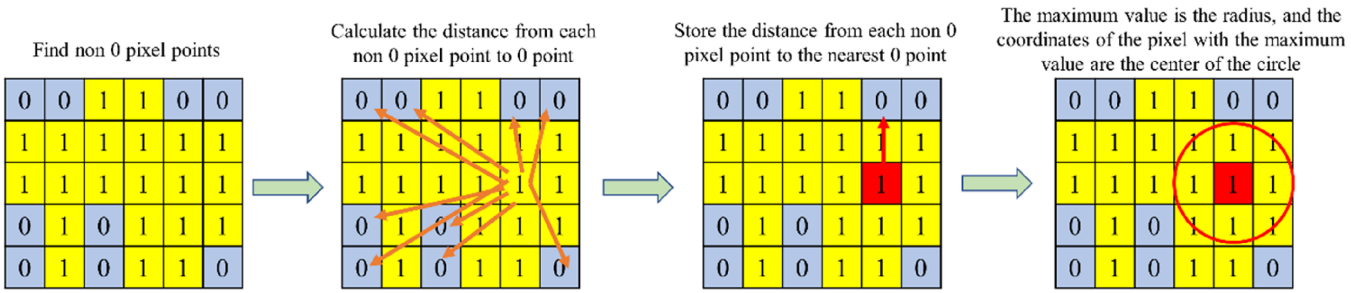

FIGURE 4 The pipeline of the distance transformation algorithm.

interested tissue region contributes more in the loss function. In addition, if x_i is predict as background but supposed to be the tissue, $\log(P(x_i))$ will be unfortunately small. However, the coefficient $(1-P(x_i))^{\gamma}$ is enlarged to highlight the misclassified pixel x_i. On the contrary, if x_i is correctly classified as tissue, $P(x_i)$ will approach 1 and $1-P(x_i)$ is accordingly very small. By using a large γ, the correctly classified pixels will receive less attention in the subsequent training. We combine dice loss function and focal loss function to construct the final DF_Loss:

$$\mathrm{DF\_Loss} = 1 - \frac{2|X^p \bigcap Y^g|}{|X^p| + |Y^g|} + \sum_{i=1}^{N} \left(\frac{-\alpha \cdot c \cdot (1-P(x_i))^{\gamma} \log(P(x_i)) - (1-\alpha)(1-c) P(x_i)^{\gamma} \log(1-P(x_i)))}{N} \right). \quad (3)$$

The VGG-SeUnet is trained by DF_Loss for 60 epochs, with a batch size of 2 and a learning rate of $5 \times e^{-4}$. The model is fitted using the Adam optimizer.

3.3 | Automatic morphology-based measurements

In order to measure the thickest subcutaneous soft tissue areas from the segmentation results, this paper establishes a morphology-based processing scheme. To avoid the interference of the background, the segmented image is converted into a binary map to facilitate further processing. In the binary map, the ROI region (tissue region) is represented by 1 and the background region is denoted by 0. A distance transformation algorithm is adopted to store the MDs between each non-zero pixel and all the zero pixels, and then retrieve the maximum value from the records. We regard the pixel where the maximum distance is retrieved as the center of the maximum inscribed circle, and draw this circle whose radius equals the maximum distance. The diameter stands for the FASSTT. The distance transformation algorithm can be interpreted by Figure 4.

The automatic measurement of FASSTT can be summarized as Algorithm 1, where p represents the position of the pixel.

The automatic measurement of FASSTT can be depicted by Figure 5. Compared with clinical manual measurements, our method has higher operational efficiency and better generalization ability.

To give the readers better insight, we conduct unit conversion from length in pixels to centimeters. The conversion depends on the number of pixels per inch (DPI):

$$L_c = L_p \times 2.54/DPI$$

where L_p denotes the length in pixels, L_c stands for the length in centimeters and 2.54 means one inch equals 2.54 centimeters. As the image resolution is 96 DPI, 1 cm can be approximated by 38 pixels in the conversion to obtain the actual subcutaneous soft tissue thickness for simplicity.

4 | EXPERIMENTS AND RESULTS

4.1 | Datasets and pre-processing

All experiments conducted in this paper used ultrasound imaging data of fetal AC from Hangzhou Women's Hospital, China. These data were collected from August 2020 to November 2021, and the segmentation boundary and

ALGORITHM 1 Automatic measurement of FASSTT

Input:

Segmented ultrasound image S

Output:

Maximum inscribed circle image and FASSTT

1: Initialize $M=$ to store the minimum distance

2: **While** there are still non-zero pixels in S **do**:

3: Select one non-zero pixel s_n

4: Initialize $D=$ to store the distance

5: **For** each zero pixel s_z **do**

6: Calculate $d_z = \left\| p_{s_n} - p_{s_z} \right\|$

7: $D = D \cup d_z$

8: **End for**

9: Retrieve the minimum value d_{min} from D

10: $M = M \cup d_{min}$

11: **End while**

12: Retrieve the maximum value d_{max} from M

13: Locate the pixel position $[x, y]$ where d_{max} emerges

14: Draw the maximum inscribed circle with center $[x, y]$ and radius d_{max}

15: FASSTT $= d_{max} \times 2$

tissue thickness in each image were annotated by two senior ultrasound chief doctors. The data have been reviewed by the Ethics Committee under the approval number 2023 Medical Ethics Review A (042). The measurements were annotated in the form of file names and the data were desensitized, as shown in Figure 6. The file names consist of a number and length in centimeters. The data set contains 135 fetal AC images and 135 segmented and annotated images in JPG and PNG formats, respectively, with a resolution of 1280 × 872 and 96 pixels per inch (DPI). Data inclusion criteria: (1) Intrauterine singleton pregnancy; (2) Chinese nationality, living in East China for more than 2 years; (3) The time between the last ultrasound examination before delivery and the delivery time is less than or equal to 3 days; (4) Definite gestational week, confirmed by fetal head-rump length measured by ultrasound examination before 13 + 6 weeks of gestation; If ultrasound CRL results were not obtained, gestational week could be confirmed by fetal BPD, HC, AC and FL by mid-trimester ultrasound examination; (5) Completion of pregnancy documentation in our hospital. Exclusion criteria: (1) Intrauterine multiple pregnancy; (2) Week of gestation unknown; (3) Presence of one or both spouses of non-Chinese nationality; (4) Fetal abnormalities suggested by ultrasound or chromosomal examination.

In order to reduce the computational complexity of the model and improve performance, the images of the dataset were normalized and converted to a resolution of 512 × 512 before they were sent to the network for

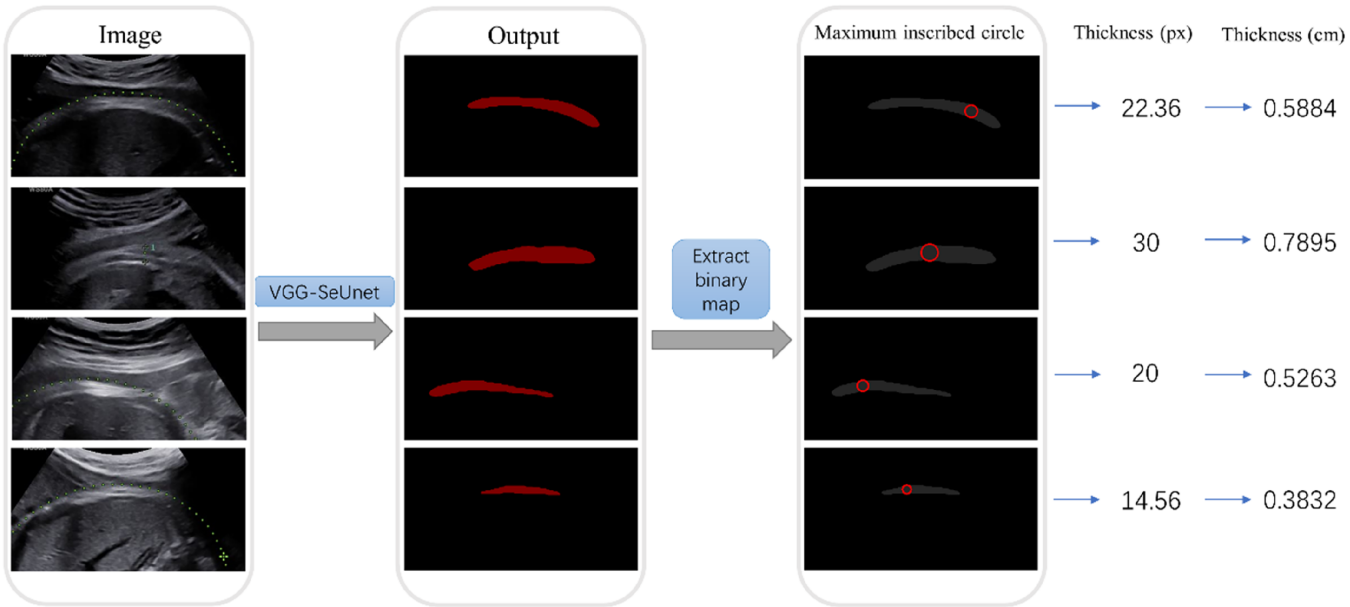

FIGURE 5 Automatic measurement process (under 96DPI image resolution).

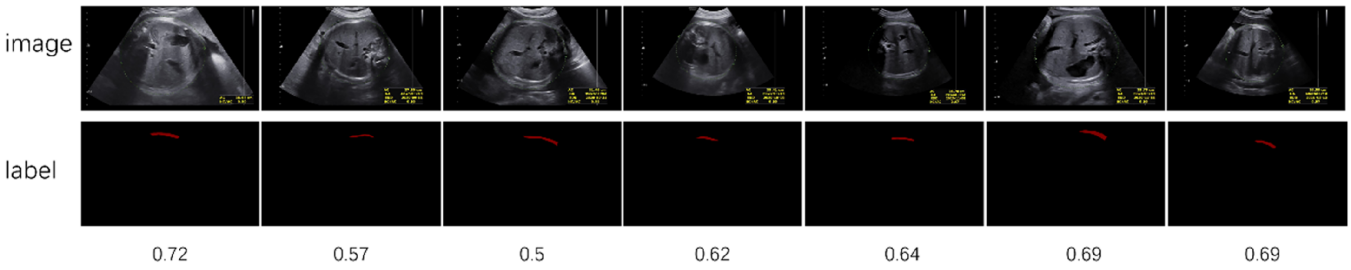

FIGURE 6 Fetal abdominal circumference ultrasound image dataset from Hangzhou Women's Hospital.

training. Note that altering the image resolution will not change the DPI because DPI denotes the number of pixels that can be rendered in each inch output of the display devices. The designated resolution meets the requirements of the model, satisfactorily preserves the structural details of the images, and significantly reduces the computational cost of the model. In terms of dataset partitioning, 20% of the images were randomly selected as the test set and 80% as the training set. Also, considering the small number of datasets, the training images are randomly enhanced by flipping, scaling, aspect warping, and color gamut transformation to avoid overfitting and enhance the generalization ability of the model.

4.2 | Evaluation criteria

In order to objectively and quantitatively measure the effectiveness of the model, seven representative evaluation metrics are used, including Average Error (AVE), Root Mean Square Error (RMSE), Dice Coefficient (DSC), Mean Intersection Ratio (mIou), Mean Pixel Accuracy (mPA), Precision and Recall. AVE evaluates the measurement results by comparing the actually measured thickness with the labeled thickness. RMSE reduces the impact of the error about individual sample to the total error to improve the objectivity. Lower value of AVE and RMSE denotes better result. The AVE and RMSE can be computed through formula (4) and (5) respectively.

$$\mathrm{AVE} = \frac{1}{n}\sum_{i=1}^{n}\left|x^i_{label} - x^i_{pred}\right|, \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x^i_{label} - x^i_{pred}\right)^2}, \tag{5}$$

where x^i_{label} denotes the truth thickness and x^i_{pred} denotes the predicted thickness. The remaining evaluation metrics are used to evaluate the segmentation result of the model, and the higher the value is, the more accurate the segmentation is. DSC can be computed as follows:

$$\mathrm{DSC} = \frac{2\times\left|y_{label}\bigcap y_{pred}\right|}{\left|y_{label}\right| + \left|y_{pred}\right|}, \tag{6}$$

where y_{label} represents the annotated ROI (tissue) region (pixel set) and y_{pred} denotes the segmented tissue region (pixel set). The principle about DSC has been explained in Section 3.2. mIou can be computed as follows:

$$\mathrm{mIou} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{TP_i}{FN_i + FP_i + TP_i}, \tag{7}$$

where k is the category ID, and category 0 denotes the background class. In formula (7), TP_i, FN_i, FP_i stands for the number of True Positive, False Negative, and False Positive samples with respect to each category. Similarly, by defining TN_i as the number of True Negative samples, we obtain the mPA:

$$\mathrm{mPA} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}. \tag{8}$$

Precision and Recall are used to evaluate the segmentation performance of all categories:

$$\mathrm{Precision} = \frac{TP}{TP + FP}, \tag{9}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN}. \tag{10}$$

4.3 | Ablation experiments

Tables 1 and 2 show the results of the ablation experiment by comparing the segmentation and measurement evaluation indicators of different module method

TABLE 1 Segmentation performance of different module configurations of VGG-SeUnet.

Method	DSC (%)	mIou (%)	mPA (%)	Precision (%)	Recall (%)
U-net	78.20%	70.30%	73.07%	**88.80%**	73.07%
VGG-Unet	86.70%	79.16%	85.02%	88.67%	85.20%
VGG-SeUnet	87.80%	79.64%	86.84%	87.65%	86.84%
VGG-SeUnet-DF_Loss	**88.90%**	**81.02%**	**90.44%**	86.38%	**90.44%**

Note: The bold values indicate the best results about each evaluation metric.

TABLE 2 Measurement performance of different module configurations of VGG-SeUnet.

Method	AVE (cm)	RMSE (cm)
U-net	0.1700	0.230
VGG-Unet	0.0870	0.100
VGG-SeUnet	0.0830	0.097
VGG-SeUnet-DF_Loss	**0.0615**	**0.081**

Note: The bold values indicate the best results about each evaluation metric.

configurations. Table 1 shows the segmentation results, while Table 2 shows the FASSTT measurement results. In these two tables, Unet means that the original Unet network structure is used and the cross-entropy loss function is used for training; VGG-Unet means that the VGG-16 network used for feature encoding is merged into Unet; VGG-SeUnet means that the channel Attention (Se module) is introduced into each jump connection in VGG-Unet; VGG-SeUnet-DF_Loss means replacing the cross-entropy loss function in VGG-SeUnet with the DF_Loss function, the best results in the table are in bold mark.

Table 1 indicates that the VGG-Unet outperforms U-net considerably with DSC increasing from 78.2% to 86.7%, and the rest segmentation metrics also rise apparently except that the Precision decreases slightly. The reason is two-fold. First, there is usually a trade-off between Precision and Recall, that is, higher Recall usually leads to lower Precision. More importantly, the lower Precision is caused by the additionally segmented soft tissue area compared with the manually labeled ground truth. The subcutaneous soft tissue is the peripheral annular region that surrounds the entire abdomen of the fetus, and we have to enlarge the segment area to guarantee that it contains the thickest location of the subcutaneous soft tissue. For the thickness measurement task, improving the performance of the Recall is more important and more critical than improving the Precision.

Table 2 shows that the AVE of VGG-Unet is reduced to 0.087 cm, which is far less than that of U-net, that is, 0.17 cm. Similarly, the RMSE decreased from 0.23 cm to 0.1 cm. Both measurement metrics are reduced by more than 50%. Except for the significant improvement of the holistic measurement result, the VGG-Unet does not produce large errors for individual images with measurements difference greater than 0.3 cm, which further demonstrates that a good encoder can effectively improve segmentation and measurement result. By incorporating the channel attention that highlights the important channels to segmentation and the DF_Loss that mitigates the sample imbalance problem into VGG-Unet, the measurement results yield further improvements. Specifically, the AVE of VGG-SeUnet-DF_Loss is reduced to 0.065 cm and the RMSE decreases to 0.081 cm, both of which are less than 0.09 cm. The ablation experimental results prove that our model has higher segmentation performance and measurement accuracy. To facilitate visual observation, the absolute values of the differences between the predict and label thicknesses in the ablation experiment are rendered into histograms and shown in Figure 7, where the horizontal axis denotes the sample ID while the vertical axis stands for the measurement error in length (cm). As VGG, channel attention, and DF-Loss are successively incorporated into the basic U-Net, the measurement error is continually decreasing and there are no longer individual samples with large errors, which prove the validity of each module.

4.4 | Comparative experiments

We compared the segmentation as well as measurement results of the proposed method with that of several representative methods including U-net, ResNet50[19]-Unet, SegNet,[20] Mobilenet-DeepLabV3,[21] LedNet,[22] ResNet50-Pspnet[23] and Segformer.[24] The segmentation comparison is shown in Table 3 and the measurement comparison is shown in Table 4, where the best results are shown in bold.

Table 3 shows that our method achieved the highest DSC (88.9%), mIou (81.02%), mPA (90.44%), and Recall (90.44%) among all the methods. Although Precision is slightly lower than ResNet50-Unet, our method achieved better holistic performance than other methods. We have

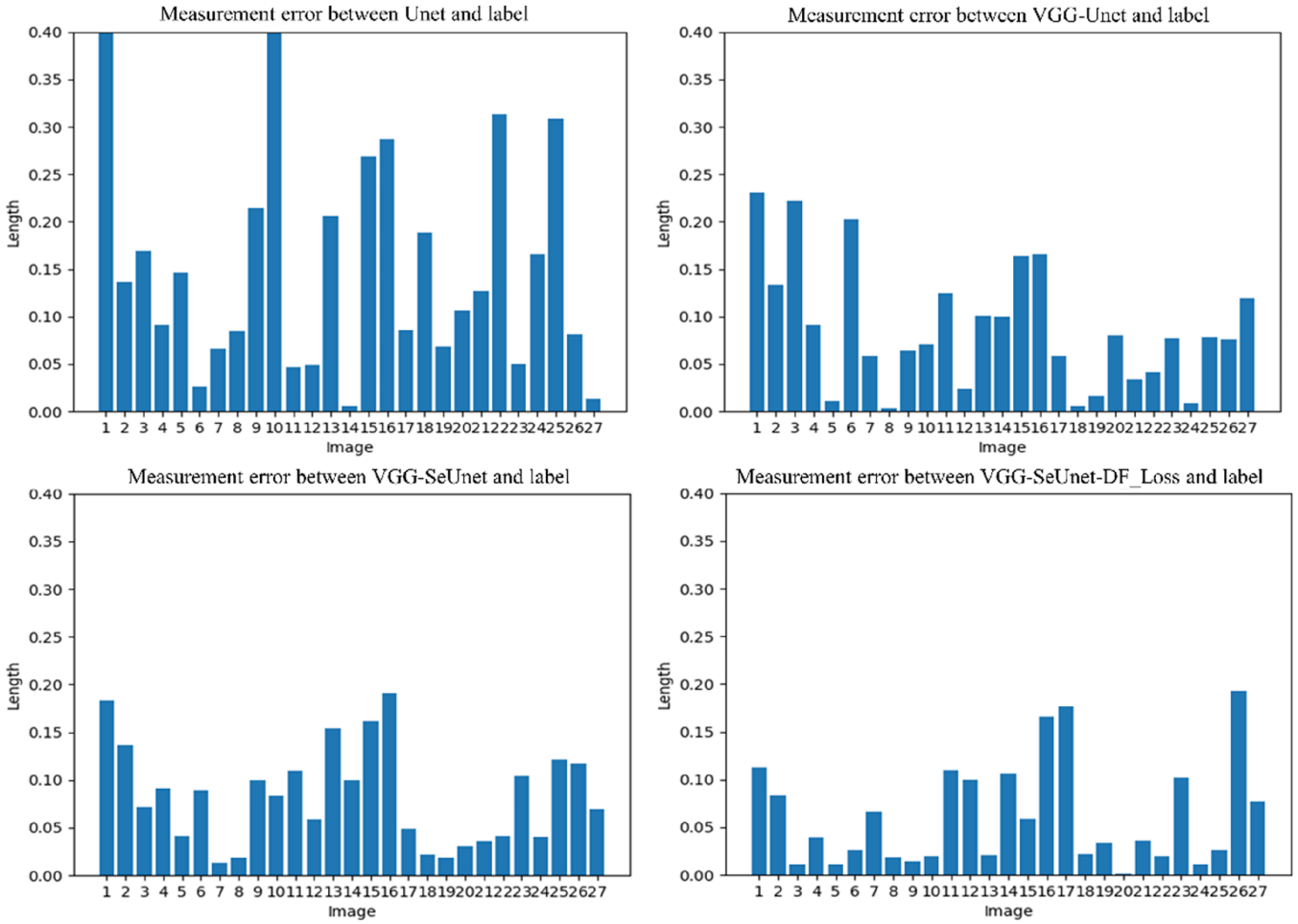

FIGURE 7 Histogram of the absolute value of the difference between the predicted and labeled thicknesses in the ablation experiment on the test set.

TABLE 3 Segmentation results of our method and several representative methods.

Method	DSC (%)	mIou (%)	mPA (%)	Precision (%)	Recall (%)
Unet[15]	78.20%	70.30%	73.07%	88.80%	73.07%
ResNet50[19]-Unet	87.70%	80.14%	86.17%	**89.2%**	86.17%
SegNet[20]	83.90%	74.90%	81.43%	85.37%	81.43%
Mobilenet-DeepLabV3[21]	82.80%	73.56%	80.71%	83.55%	80.71%
LedNet[22]	75.40%	67.21%	78.20%	73.61%	78.20%
ResNet50-Pspnet[23]	81.20%	74.85%	80.81%	86.08%	80.81%
SegFormer[24]	77.6%	69.89%	74.88%	83.42%	74.88%
VGG-SeUnet-DF_Loss	**88.9%**	**81.02%**	**90.44%**	86.38%	**90.44%**

Note: The bold values indicate the best results about each evaluation metric.

discussed the reason leading to lower Precision in Section 4.3. Table 4 shows that our method had the smallest AVE (0.0615 cm) and RMSE (0.081 cm), and no samples with large errors were observed. This proves that our method achieves superior measurement accuracy to other methods.

To intuitively show the segmentation results and the positions where the FASSTTs were measured, we visualized the prediction results by our method and several representative methods in Figure 8, where the green circle indicates the maximum inscribed circle under the doctor's segmentation label (the diameter is the labeled

TABLE 4 Measurement results of our method and several representative methods.

Method	AVE (cm)	RMSE (cm)
Unet[15]	0.1700	0.2300
ResNet50[19]-Unet	0.0849	0.1100
SegNet[20]	0.0930	0.1200
Mobilenet-DeepLabV3[21]	0.1800	0.2240
LedNet[22]	0.3000	0.3500
ResNet50-Pspnet[23]	0.1100	0.1400
SegFormer[24]	0.1669	0.2245
VGG-SeUnet-DF_Loss	**0.0615**	**0.0810**

Note: The bold values indicate the best results about each evaluation metric.

thickness), and red circles are the maximum inscribed circles drawn from the predicted ROI area. As the segmentation area to be measured is small, we have enlarged and cropped the prediction results and the original images to facilitate comparison between different methods.

The first row shows the original image, the second row shows the annotated image and the corresponding positions of the maximum inscribed circles, and the third to ninth rows show the subcutaneous soft tissue segmentation results of different methods, and the positions where tissue thickness was measured. The segmentation result obtained by our method is most similar to the annotated image with smooth edges. The locations of the maximum inscribed circles derived from our method are also closest to that of the labeled images.

Moreover, Figure 8 intuitively explains why our method yields lower Precisions in Tables 1 and 3. In Figure 8, the segmented area (rendered in dark gray) by all the involved methods is a section of subcutaneous soft tissue, which is usually much bigger than the ground truth region annotated by the doctor. This is to include the thickest area of the subcutaneous soft tissue as much as possible. As a consequence, the segmentation results inevitably decrease the Precision.

5 | DISCUSSION

Based on an encoder-decoder paradigm, this paper proposes a new U-shaped attention network, VGG-SeUnet, for automatic segmentation of the subcutaneous soft tissue region of the fetal abdominal and a morphology-based measurement algorithm to measure FASSTT. The VGG-16 network in the encoder part improves the multi-level feature extraction capability of the model. A channel attention module is used between the encoder and decoder for re-weighting the feature maps and highlighting the channels with key information while suppressing the rest for better feature fusion. By performing convolution dimension reduction and LeakyReLU activation in the classifier before outputting the number of classification channels, the model complexity is reduced and the segmentation performance is improved. In addition, the DF_Loss function exploits both dice loss and focal loss to solve the problem of sample imbalance between foreground and background regions. Numerous experiments have shown that our method can effectively solve the segmentation problem with difficulties including complex surrounding tissues, unclear edges, and unbalanced sample distribution in the subcutaneous soft tissue region.

Based on the segmented ROI region, this paper proposes a morphological method to automatic measurement of FASSTT. The key to this method is a Distance Transform algorithm, which allows for simple and efficient localization of the thickest position for accurate measurement. Experiments show that our method not only distinguishes well between subcutaneous soft tissue areas and background but also outperforms existing algorithms in terms of DSC, mIou, mPA, Recall, AVE, and RMSE. Our method achieves the lowest measurement error, which is below 0.09 cm, providing the best results in comprehensive assessment. Due to the limited training samples and the lack of publicly available datasets for fetal subcutaneous soft tissue thickness measurements, we had to rely on data augmentation to increase the number of training images, which is a limitation of this paper. The proposed model would have performed better if it could have been trained on a larger dataset.

6 | CONCLUSION

In this paper, we combine the VGG-SeUnet network, DF_Loss loss function, and a morphology-based measurement algorithm to propose an automatic approach to FASSTT measurement. Our method does not consume a large amount of computational resources. The VGG-SeUnet network is trained in an end-to-end manner without the need for manual feature extraction and possesses superior segmentation performance compared with existing state-of-the-art networks. The morphology-based measurement algorithm subsequently yields promising measurement results. The proposed approach enables imaging physicians to automate the measurement of FASSTT, so as to assist physicians in the subsequent assessment of fetal development and nutritional status.

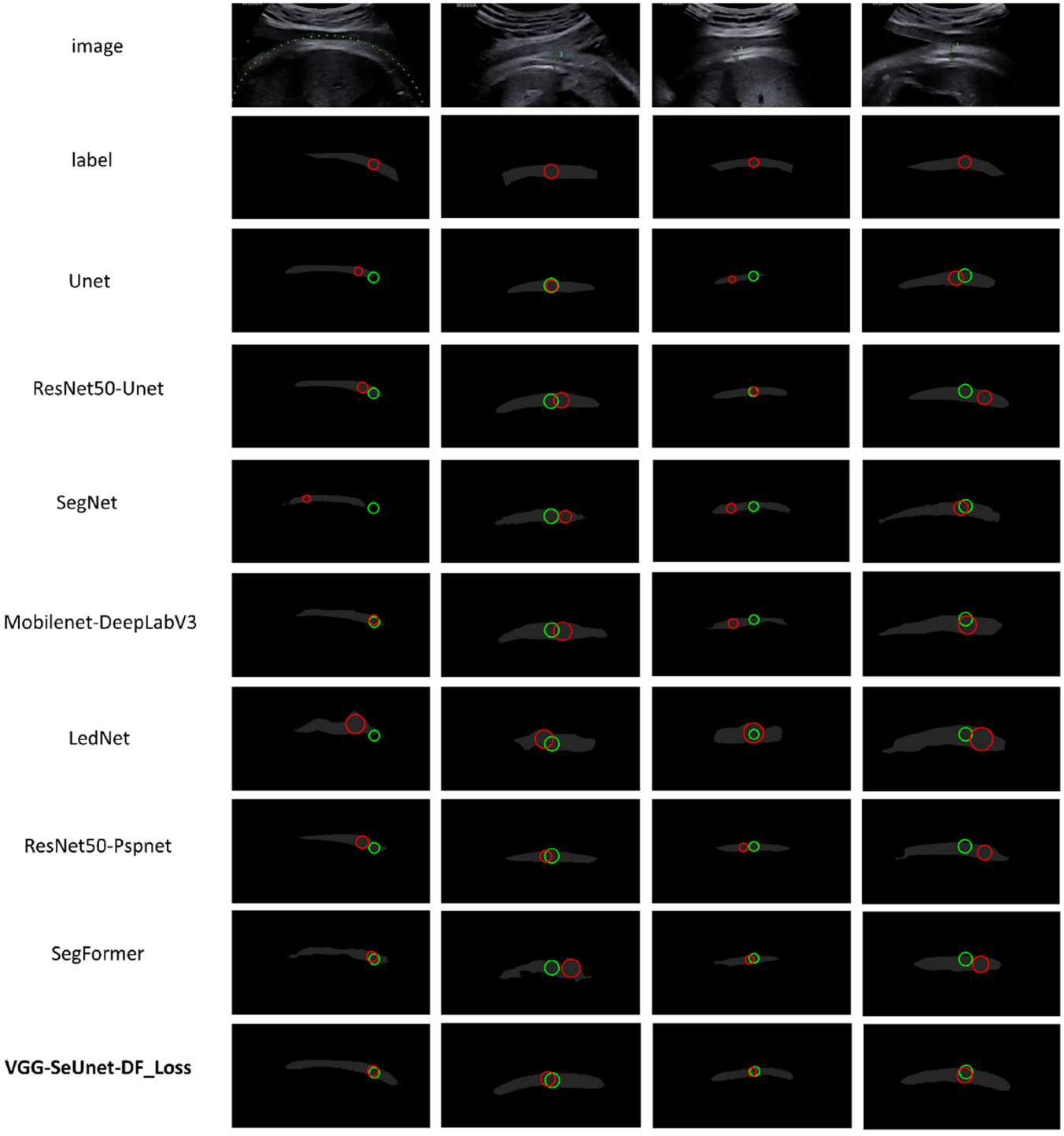

FIGURE 8 Visualization of the segmentation results yielded by different methods and the maximum inscribed circle subsequently derived from the segmentation results. Green circles indicate the positions where the FASSTT was annotated by doctors (the diameter is the labeled thickness), and red circles are the maximum inscribed circles drawn from the predicted ROI area.

AUTHOR CONTRIBUTIONS

Zhenming Yuan and Jian Zhang: Thesis guidance and revision; Tianhao Xu: Programming and initial draft writing; Cheng Yu and Xiaojun Ye: Data collection and annotation. All authors reviewed the manuscript.

FUNDING INFORMATION

This paper is supported by the National Natural Science Foundation of China under Grant No. 61972361; Primary Research and Development Plan of Zhejiang Province in China under Grant No. 2020C03107.

CONFLICT OF INTEREST STATEMENT

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service, and company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Tianhao Xu https://orcid.org/0000-0001-5570-5321
Cheng Yu https://orcid.org/0000-0001-8101-8330
Xiaojun Ye https://orcid.org/0000-0002-2481-3351
Jian Zhang https://orcid.org/0000-0001-6478-9192

REFERENCES

1. Xie X, Kong B, Duan T. *Obstetrics and Gynecology*. 9th ed. People's Medical Publishing House; 2019:53.
2. Pascal A, Govaert P, Oostra A, Naulaers G, Ortibus E, Van Den Broeck C. Neurodevelopmental outcome in very preterm and very-low-birthweight infants born over the past decade: a meta-analytic review. *Dev Med Child Neurol*. 2018;60(4): 342-355.
3. Song Q, Luo H. The important significance of applying ultrasound indicators to evaluate fetal growth and development. *Chin J Med Ultrasound*. 2021;18(8):733-736.
4. Lu Q, Sun Y. Clinical observation of ultrasound measurement of fetal abdominal subcutaneous tissue thickness for predicting fetal weight. *Chin J Pract Gynecol Obstetr*. 2007; 06:450-451.
5. Köşüş N, Köşüş A. Can fetal abdominal visceral adipose tissue and subcutaneous fat thickness Be used for correct estimation of fetal weight? A preliminary study. *J Obstet Gynaecol*. 2019; 39(5):594-600.
6. Khalifa EA, Hassanein SA, Eid HH. Ultrasound measurement of fetal abdominal subcutaneous tissue thickness As a predictor of large versus small fetuses for gestational age. *Egypt J Radiol Nucl Med*. 2019;50(1):80.
7. Oghli MG, Shabanzadeh A, Moradi S, et al. Automatic fetal biometry prediction using a novel deep convolutional network architecture. *Phys Med*. 2021;88:127-137.
8. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proc IEEE Confer Comput Vis Pattern Recogn*. 2018;7132-7141.
9. Zeng Y, Tsui PH, 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*. 2021; 34(1):134-148.
10. Zeng W, Luo J, Cheng J, Lu Y. Efficient fetal ultrasound image segmentation for automatic head circumference measurement using a lightweight deep convolutional neural network. *Med Phys*. 2022;49(8):5081-5092.
11. Oghli MG, Moradi S, Sirjani N, et al. Automatic measurement of fetal head biometry from ultrasound images using deep neural networks. Proceedings of the 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference, 2020:1–3.
12. Zhu F, Liu M, Wang F, Qiu D, Li R, Dai C. Automatic measurement of fetal femur length in ultrasound images: a comparison of random Forest regression model and SegNet. *Math Biosci Eng*. 2021;18(6):7790-7805.
13. Kim B, Kim KC, Park Y, Kwon JY, Jang J, Seo JK. Machine-learning-based automatic identification of fetal abdominal circumference from ultrasound images. *Physiol Meas*. 2018;39(10):105007.
14. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Proceedings of International Conference on Learning Representations 2015.
15. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234–241.
16. Shu X, Gu Y, Zhang X, Hu C, Cheng K. FCRB U-net: a novel fully connected residual block U-net for fetal cerebellum ultrasound image segmentation. *Comput Biol Med*. 2022;148:105693.
17. Milletari F, Navab N, Ahmadi SA. V-net: fully convolutional neural networks for volumetric medical image segmentation. Proceedings of the Fourth International Conference on 3D Vision (3DV), 2016: 565–571.
18. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980–2988.
19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778.
20. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(2):2481-2495.
21. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with Atrous separable convolution for semantic image segmentation. Proceedings of the European Conference on Computer Vision, 2018: 801–818.
22. Wang Y, Zhou Q, Liu J, Xiong J, Gao G, Wu X, Latecki JL. Lednet: a lightweight encoder-decoder network for real-time semantic segmentation. Proceedings of the 2019 IEEE International Conference on Image Processing. 2019. 1860–1864.
23. Zhao H, Shi J, Qi X, Wang X Jia J. Pyramid scene parsing network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881–2890.
24. Xie E, Wang W, Yu Z, Anandkumar A, Alvarez JM, Luo P. SegFormer: simple and efficient design for semantic segmentation with transformers[J]. *Adv Neural Inform Process Syst*. 2021;34: 12077-12090.

How to cite this article: Yuan Z, Xu T, Yu C, Ye X, Zhang J. Automatic measurement of fetal abdomen subcutaneous soft tissue thickness from ultrasound image based on a U-shaped attention network with morphological method. *Int J Imaging Syst Technol*. 2024;34(1):e23031. doi:10.1002/ima.23031