



Rapid and automatic assessment of early gestational age using computer vision and biometric measurements based on ultrasound video

Yuanyuan Pei^{1#^}, Wenjing Gao^{1#^}, Longjiang E^{1^}, Changpin Dai^{2^}, Jin Han^{3^}, Haiyu Wang^{2^}, Huiying Liang^{1,4,5^}

¹Clinical Data Center of Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China; ²Department of Ultrasonography, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China; ³Prenatal Diagnosis Center of Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China; ⁴Clinical Data Center, Guangdong Provincial People's Hospital/Guangdong Academy of Medical Sciences, Guangzhou, China; ⁵Guangdong Provincial Children's Medical Research Center, Guangzhou, China

Contributions: (I) Conception and design: Y Pei; (II) Administrative support: H Liang, H Wang, J Han; (III) Provision of study materials or patients: C Dai, H Wang; (IV) Collection and assembly of data: Y Pei, W Gao, L E; (V) Data analysis and interpretation: Y Pei; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Haiyu Wang. Department of Ultrasonography, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, 9 Jinsui Road, Guangzhou 510623, China. Email: haiyuwang0708@gmail.com; Huiying Liang. Research Department of Medical Sciences of Guangdong Provincial People's Hospital, 106 Zhongshan 2nd Road, Guangzhou 510080, China. Email: lianghuiying@hotmail.com.

Background: Early gestational age (GA) assessment using ultrasound is a routine and frequent examination performed in hospitals whereby clinicians manually measure the size of the gestational sac using ultrasound and calculate GA. However, the error is often substantial, and the process is laborious. To overcome these challenges, we propose a new system to assess early GA using a new end-to-end computer vision system and a new biometric measurement method based on ultrasound video.

Methods: In this retrospective study, a new system was provided. B-ultrasound videos were first decomposed into two-dimensional (2D) images, and the contours of the gestational sac were extracted and drawn. The maximum length and short diameter of the gestational sac were then automatically measured and GA was calculated using the Hellman formula. Finally, through human-machine comparison, the clinicians' assessment errors were analyzed by SPSS 26.

Results: In this study, 29,829 2D images of 191 B-ultrasound videos were evaluated using the new system. Clinicians usually require 15–20 min to complete assessments of GA, whereas with the new system assessments can be completed in approximately 30 s. Moreover, a human-machine comparison showed that the system helped intermediate skills clinicians improve their relative diagnostic error by 13.45% with an absolute error of 7 days. In addition, the new system was used to identify other lesions in the uterus and measure their size as a “sanity check”.

Conclusions: The proposed new system is a practical, reproducible, and reliable approach for assessing early GA.

Keywords: Computer vision; gestational age (GA); gestational sac; biometric measurement; B-ultrasound

[^] ORCID: Yuanyuan Pei, 0000-0002-9283-5958; Wenjing Gao, 0000-0001-5138-0785; Longjiang E, 0000-0002-8870-7750; Changping Dai, 0000-0002-5201-600X; Jin Han, 0000-0002-6005-1905; Haiyu Wang, 0000-0002-0240-7520; Huiying Liang, 0000-0002-9987-8002.

Submitted Aug 23, 2021. Accepted for publication Dec 09, 2021; Published online: 04 Jan 2022.

doi: 10.21037/qims-21-837

View this article at: <https://dx.doi.org/10.21037/qims-21-837>

Introduction

In the first trimester of pregnancy, determining gestational age (GA) is the basis of obstetric diagnosis and treatment, and accurate calculation of the expected date of delivery helps determine the timing of prenatal examinations and prenatal diagnoses (1,2). A common method to predict GA is by manually measuring the maximum length and short diameter of the gestational sac using ultrasound (3,4). However, a B-ultrasound video is dynamic, and the gestational sac is three-dimensional (3D). Individual clinicians observe gestational sacs from different angles, resulting in various sizes. Each time, clinicians select the largest two-dimensional (2D) image of the gestational sac in the video based on personal experience and then use the electronic ruler of the B-ultrasound machine to manually measure the length and short diameter of the gestational sac to calculate GA. This is a time-consuming task that requires skill and experience and is prone to human error. Some pregnant women receive different estimates of GA in different hospitals on the same day.

In recent years, there has been considerable research into the assessment of GA in the first trimester; however, to the best of our knowledge, a fast and accurate clinically applicable method has not yet been developed. In 2012, Zhang *et al.* (5) evaluated the use of automatic standard plane selection and biological measurements of the early pregnancy sac. In their study, a square frame was used to measure the gestational sac, and the measurement contained large errors (5). In 2015, Chen *et al.* (6) proposed an automatic detection method for a fetal ultrasound standard plane based on the knowledge transfer recurrent neural network; however, their results contained substantial errors. In 2016, Ibrahim *et al.* (7) proposed a method for the automatic segmentation and measurement of the pregnancy sac; however, this was based on B-ultrasound static pictures instead of videos, and the implementation process was cumbersome and inconvenient for clinical application. In 2019, Kim *et al.* (8) used machine learning ultrasound images to study fetal head biometrics, but they did not study the smaller gestational sac.

Computer vision is an effective technology for simplifying the process of medical imaging (9-12). Herein, we propose a new biometric measurement method that

will provide improved accuracy in the measurement of medical images. We describe an innovative technique to automatically draw the outline and measure the maximum length and short diameter of the gestational sac to assess GA. We present the following article in accordance with the Standards for Reporting of Diagnostic Accuracy Studies (STARD) reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-21-837/rc>).

Methods

Video data (n=191) were obtained from the Guangzhou Women and Children Medical Center, China, for ultrasound examinations conducted between January and December 2018. A series of experimental steps was conducted on the B-ultrasound videos from the first trimester of pregnancy (between 4.6 and 11 weeks). Notably, in early pregnancy, the gestational sac forms at >4.6 weeks GA and then disappears, becoming the placenta at >11 weeks. The experimental steps were as follows: (I) decomposition of the B-ultrasound video into a 2D plane image in the order of the original frames using OpenCV4 (<https://opencv.org/opencv-4-0/>); (II) filtering of the 2D plane image, followed by detection of the edge of the gestational sac and extraction of its outline using computer vision; (III) automatic measurement of the length and short diameter of the gestational sac using a new biometric measurement method; (IV) calculation of the age of the gestational sac according to the Hellman formula (13); and (V) comparison of the diagnostic capabilities of this new system with the diagnostic capabilities of clinicians.

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The protocol was reviewed and approved by the Ethics Committee of the Guangzhou Women and Children's Medical Centre (Scientific Research Ethics Committee permission No. GO-2016-017). Written informed consent was provided by all participants at the time of their initial hospital visit.

Study design

In this study, a new end-to-end computer vision system was proposed. Using the OpenCV library of programming

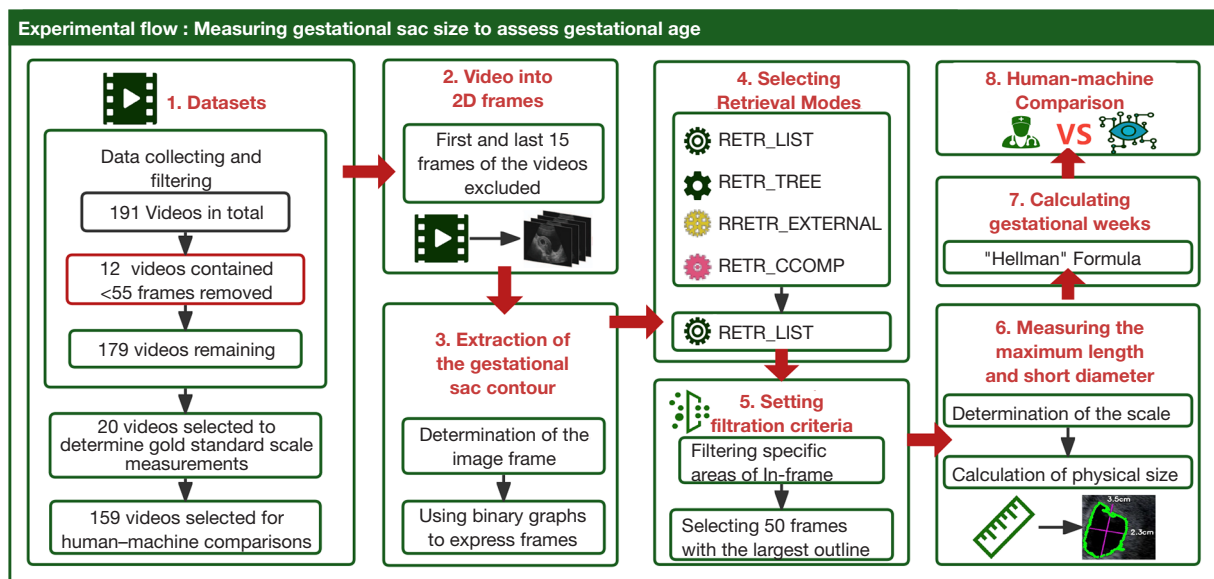


Figure 1 Experimental flow.

Table 1 Contour retrieval modes and definitions

| Retrieval method | Retrieval mode | Definition |
|------------------|----------------|---|
| 1 | RETR_LIST | Subordinate relationships between contours are not established, and all contours belong to the same level |
| 2 | RETR_TREE | Hierarchical affiliations of the profiles completely established |
| 3 | RETR_EXTERNAL | Only the highest-level contour (i.e., level 0) is found |
| 4 | RETR_CCOMP | All contours are divided into 2 levels: the outer layer and the inner layer |

functions, the length and short diameter of the gestational sac in B-ultrasound videos were measured quickly and accurately to predict GA. Assessment results were compared between the new system and those made by intermediate skills clinicians. The experimental process was divided into 8 stages (*Figure 1*).

For a description of retrieval modes, see *Table 1*.

Datasets

The B-Ultrasound videos of women in early pregnancy who were examined at Guangzhou Women and Children's Medical Centre between January 2018 and December 2018 were analyzed retrospectively. The dataset included 191 ultrasound videos of GAs between 4.6 and 11 weeks, each acquired at the same magnification. The original video had 768×576 pixels per frame. The scans were collected by a team of 5 experienced sonographers on a Mindray B-mode

ultrasound apparatus (Mindray, Shenzhen, China) with a 2–5 μHz probe.

When the videos were decomposed into 2D plane images, 12 videos contained <55 frames because the measurement time was too short (2 s) to accurately obtain the maximum length and short diameter of the gestational sac; hence, these videos were filtered. A total of 20 videos were used to determine the “gold standard” (as a reliable reference standard) of the scale by 2 senior skills clinicians (Prof. HYW and CPD), and the remaining 159 videos were used for human-machine comparison tests.

Experimental environment

All experimental steps were performed on a personal computer (Windows 7 Home Edition, 64-bit operating system) using PyCharm Professional 2020.3 (JetBrains, Forster City, CA, USA) and OpenCV4 software. The

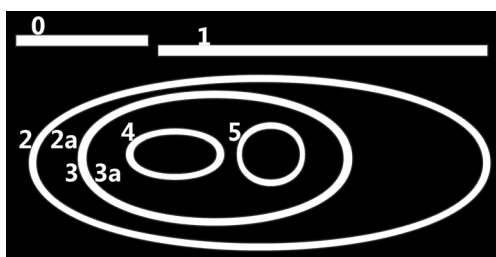


Figure 2 Sample diagram of contour outlines and hierarchy.

personal computer was not equipped with an NVIDIA graphics card.

Experimental steps

Gold standard scale

To determine the gold standard scale, the number of pixels per centimeter in 2D video images was determined by 2 senior skills clinicians (Prof. HYW and CPD) with >15 years' experience. The system provided a pixel-level accurate measurement of the maximum length and short diameter of the gestational sac, and the clinicians' measurement errors come from comparisons with this system. The GA was estimated by the Hellman formula, widely used in clinical practice, by both the system and the clinicians. All B-ultrasound video data were scanned correctly and acquired at the same magnification.

Decomposing the B-ultrasound video into a single-frame plane image

The B-ultrasound videos collected in the experiment contained 25 frames/s. Usually, B-ultrasound videos of an early pregnancy examination have approximately 200–600 frames of 2D plane images. In the present study, OpenCV framework programming was used to achieve an end-to-end output from the video to the 2D plane image. The original video with 768×576 pixels was used as the input, then decomposed into 2D images by the system and read in order. Finally, a 2D image of 768×576 pixels, the same size as the video, was output and stored.

Extracting the contour of the gestational sac area on the 2D decomposed video image

The uppermost and lowermost rows and the leftmost and rightmost columns of the 2D image formed the picture frame. The width and the height of the image were 768 and 576 pixels, respectively; as such, the frame composition of

the image comprised pixels in row 0, row 575, column 0, and column 767.

Converting imported images into greyscale images and rendering binary images

Pixels with greyscale values of 0 and 1 were called 0-pixels and 1-pixels, respectively; (i, j) represented a pixel in the i -th row and j -th column in the image and f_{ij} represented the greyscale value of pixel (i, j) . The binarization of an image, F , was expressed as $F=(f_{ij})$.

Determining the contours and levels in the graphics

Using the eight contours in *Figure 2* as an example, contours 2 and 2a represent the outer and inner layers, respectively, as do contours 3 and 3a. From *Figure 2*, contours 0, 1, and 2 are the outermost contours and these were all designated at the same contour level, namely level 0. Contour 2a is a sub-contour of contour 2, which is the parent contour of 2a. Contour 2a was designated as level 1. Similarly, as a sub-contour of 2a, contour 3 was designated level 2; contour 3a, as a sub-contour of 3, was designated level 4, and so on (14).

Using the cv.findContours function to select the image contour retrieval method

The OpenCV function determined contours with the same level and subordinate relationships between the contours. There were four contour retrieval methods (*Table 1*) in this study (15).

The focus of the present study was to decompose the 2D plane image from the B-ultrasound video and extract the outline of the gestational sac in the 2D image. The hierarchical affiliations between the outlines did not need to be established. Therefore, RETR_LIST was selected as the retrieval mode in the present study. In addition, the cv.RETR_LIST mode was employed in OpenCV because of its superior and stable performance.

Filtering in the contour extraction process

In every 2D plane image decomposed in the B-ultrasound video, the cv.findContours function automatically extracted all contours. Accordingly, the following filtering conditions were established:

- (I) Contours were extracted from reasonable areas within the picture frame; the upper, lower, left, and right edge areas of the image were excluded.
- (II) According to the comparison table of gestational sac size and GA, contours that were too large or too small for extraction, namely those smaller than

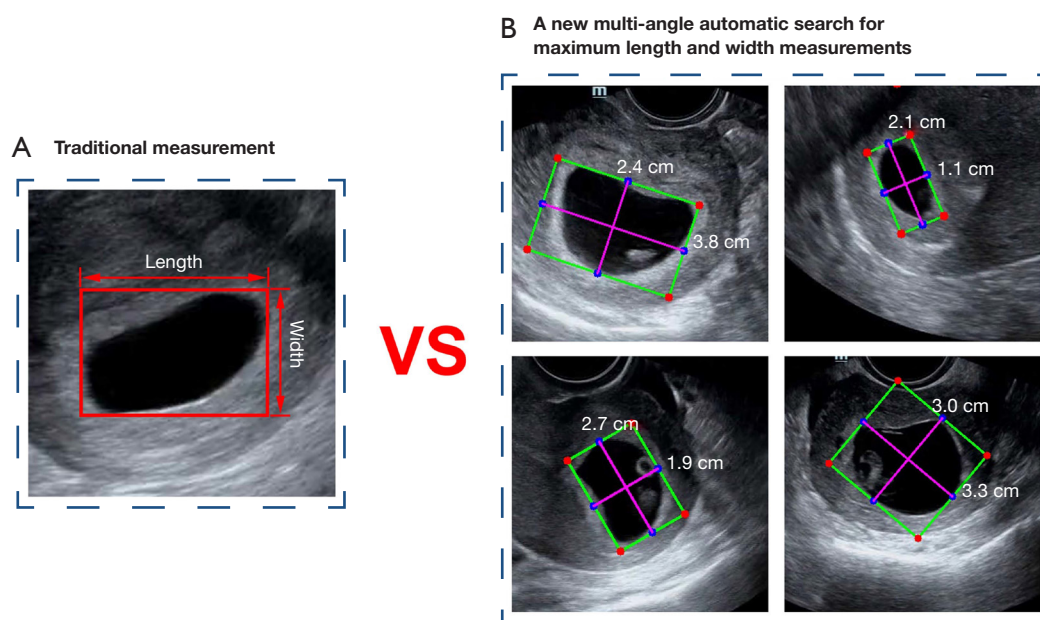


Figure 3 Comparison of traditional and new measurement method. (A) Traditional method for clinician measurements of the length and width of a gestational sac. (B) New biometric measurement method. The center point of the contour is determined using a computer program, and then minimum rectangles of the gestational sac are measured at different angles to determine the maximum long diameter and maximum short diameter.

0.6 cm × 0.8 cm or larger than 4.6 cm × 3.7 cm, were filtered.

- (III) During the video capture process with the B-ultrasound probe, the 2D plane images captured in the first and last short periods were blurred, resulting in magnification errors; therefore, the first and last 15 frames in the video were excluded.
- (IV) The largest contour in the top 50 2D images was selected and output. OpenCV extracted the contours of other objects, in addition to those of the gestational sac, such as a bleeding mass or other cysts near the uterus. We extracted contours from the 2D images decomposed from the same video and sorted them from largest to smallest, with the largest 50 contours selected as the output. This facilitated the clinician selecting the largest gestational sac and provided a basis for comprehensive differential diagnosis of other lesion areas.

Automatic measurement of the maximum length and shortest diameter of gestational sacs by scale

A new effective biometric measurement method was used. First, the number of pixels per centimeter in the 2D plane image was determined. This scale is then used to convert the number of pixels of different gestational sacs into

physical sizes.

This scale was determined to be the “gold standard” by two senior skills clinicians (Prof. HYW and CPD), each with >15 years’ experience. The scale was determined by measuring 20 gestational sacs using the electronic ruler on the B-ultrasound machine and obtaining a mean value. All videos were collected using the same method and magnification. Finally, the scale was defined as 49.4 pixels corresponding to 1 cm.

In practice, clinicians used the electronic ruler on the B-ultrasound machine according to the contour of the gestational sac to measure the largest long diameter in the horizontal direction and the largest short diameter in the vertical direction (*Figure 3A*). However, using our new measurement method, we determined the center point of the contour through a computer program and then obtained the minimum rectangle of the gestational sac from different angles to determine the maximum long diameter and maximum short diameter (*Figure 3B*). This new measurement method is more accurate than that traditionally used by clinicians.

Using the Hellman formula to determine GA according to the size of the gestational sac

To determine GA according to the size of the gestational

sac, the mean inner diameter of the gestational sac must first be determined (Eq. [2]):

$$\text{Mean inner diameter} = \frac{\text{maximum diameter} + \text{minimum diameter}}{2} \quad [1]$$

Then, GA is calculated according to the Hellman formula as follows:

$$GA(\text{weeks}) = \frac{\text{mean inner diameter}(\text{cm}) + 2.543}{0.7} \quad [2]$$

Finally, The GA obtained is automatically marked on the 2D image by our program based on OPenCV.

Human-machine comparisons

We asked five intermediate skill clinicians (with the same duration of service, experience, and ability) to assess GA using the same 159 B-ultrasound videos, and their results were compared with those obtained using the new system. Through human-machine comparison, the absolute error and measurement error of the intermediate skills clinicians and the system assessments were then statistically analyzed by SPSS version 26 (IBM Corp., Chicago, IL, USA).

Results

Using computer vision and the new biological measurement method, automatic measurements of the maximum long diameter and short diameter of the gestational sac were obtained and used to quickly assess GA. Using OpenCV, the series of complicated experimental steps was programmed on an ordinary personal computer, and the assessment of GA were obtained. The results of each of the steps are shown in *Figure 4*.

Under the condition of the same magnification of all B-ultrasound videos, the GA calculated by measuring the maximum length and short diameter of the gestational sac based on the system was considered correct. By comparing the assessments made by five intermediate skills clinicians with that of the new system, we obtained the absolute error and gestational week assessment error (*Table 2*).

One-sample statistics and independent sample *t*-tests were performed on the data in *Table 2* using SPSS version 26 as shown in *Tables 3,4*.

Based on the results of one-sample statistics (*Table 3*) and independent sample *t*-tests (*Table 4*), which indicated a relative error in the assessment of GA by intermediate skills clinicians of 11.93–14.98% (mean 13.45%) (16) and a mean of absolute error of 1.00 week, it would appear that the new

system has clinical value.

The relationship between GA and the relative error of the clinicians' assessments is shown in *Figure 5A*. As can be seen in *Figure 5B*, the relative error of clinicians' assessment of GA exceeded 30% in 8.18% of cases, with most relative errors within 25%. Furthermore, the absolute error measured by clinicians exceeded 2 weeks in 6.92% of cases, but was mostly concentrated in the range 0–2 weeks (*Figure 5C*). There were extreme cases; specifically, the absolute error exceeded 3 weeks when the actual GA was 8–10 weeks. As shown in *Figure 5A–5C*, the distribution of relative errors for clinicians assessing GA at 4.6–10 weeks was uniform, so there was no significant correlation between relative errors and GA. In addition, the red curves in *Figure 5D,5E* indicate the normal distribution curve of the probability density, showing that the relative errors are concentrated in the range of 0–20% (17,18). The mean value of the relative error was 13.45%, and the absolute error was approximately 1 week, indicating that there would be a certain deviation due only to clinicians' manual measurement. The relative error in *Figure 5D* exhibits a right-skewed distribution. The probability of a relative error >25% was low. The probability of relative errors > 50% was 0 (*Figure 5D*). The right-skewed distribution of absolute error in *Figure 5E* means that the probability of the absolute error always exceeded 0.

In the contour extraction stage, the system extracted areas that were not gestational sacs, such as bleeding or cysts near the uterus, which provides more reference information for the clinicians' diagnosis, and helps clinicians to obtain a complete diagnostic report (19,20). In addition, filtering and outputting 50 valid 2D images from a video containing several hundred frames will improve the clinicians' efficiency (*Figure 6*). The long and short diameters and dimensions are the system output results, as shown in *Figure 6*.

Discussion

Any guidance for pregnant women and their subsequent diagnosis and treatment must be based on GA (21). Calculation errors related to GA affect follow-up prenatal screening. Further, accurate calculation of GA is a key factor in determining pregnancy course and outcome. Herein, we have proposed a computer vision system for estimating GA by measuring biological characteristics and selecting the largest gestational sac measurement. This method was shown to be robust, efficient, and accurate. After filtering, a group of 50 2D pictures marked with the maximum length and shortest diameter of the gestational sac were used to calculate

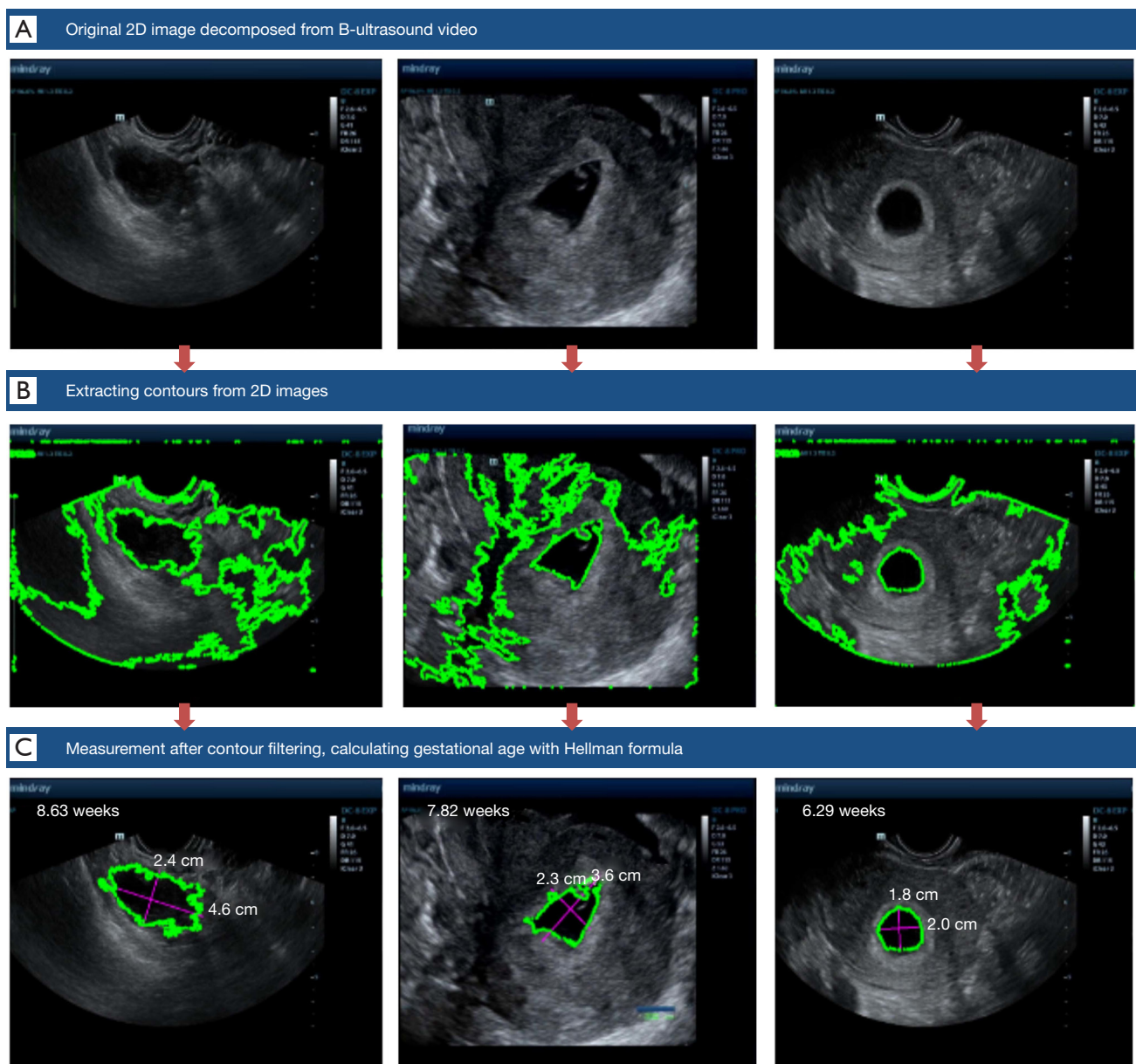


Figure 4 Results of each of the steps in the new biometric measurement method based on ultrasound video. (A) An original 2D image decomposed from the B-ultrasound video. (B) Extracting contours from the 2D images. (C) Measurement of the gestational sac after contour filtering and calculation of GA using the Hellman formula. 2D, two-dimensional; GA, gestational age.

GA, and the size of other lesion areas were obtained as the output, which reduced the complicated process of manual screening, measurement, and calculation, and facilitated clinicians' rapid and accurate estimation of GA.

It was difficult for clinicians to select the largest pregnancy sac from a B-ultrasound for measurement based on human vision and memory. We undertook human–

machine comparisons. The new system can help clinicians correct the evaluation errors, particularly when their results had large errors.

Although the method behind the new system was complicated and contained many development steps, the new system is faster and more accurate than existing methods. Importantly, all functions in the new system can

Table 2 Human-machine comparisons of the assessment of gestational age

| SN | GA (weeks) | | Relative error (%) |
|----|------------|---------|--------------------|
| | Human | Machine | |
| 1 | 7 | 6.6 | 6.06 |
| 2 | 7.3 | 7.5 | 2.67 |
| 3 | 6.3 | 6.7 | 5.97 |
| 4 | 7 | 6 | 16.67 |
| 5 | 7.3 | 7.7 | 5.19 |
| 6 | 6.3 | 6.2 | 1.61 |
| 7 | 4.3 | 5.4 | 20.37 |
| 8 | 6.3 | 9.9 | 36.36 |
| 9 | 5.5 | 6 | 8.33 |
| 10 | 7.3 | 6.5 | 12.31 |
| 11 | 8.5 | 10 | 15.00 |
| 12 | 10 | 7.5 | 33.33 |
| 13 | 5.5 | 8 | 31.25 |
| 14 | 9.3 | 8.5 | 9.41 |
| 15 | 8.3 | 8.8 | 5.68 |
| 16 | 8.3 | 7.7 | 7.79 |
| 17 | 6.7 | 5.7 | 17.54 |
| 18 | 8.3 | 8.3 | 0.00 |
| 19 | 6 | 5.3 | 13.21 |
| 20 | 6.3 | 5.7 | 10.53 |
| 21 | 6.5 | 8.2 | 20.73 |
| 22 | 9.3 | 9.7 | 4.12 |
| 23 | 9.5 | 9.8 | 3.06 |
| 24 | 9.3 | 8.5 | 9.41 |
| 25 | 8.3 | 8 | 3.75 |
| 26 | 8.3 | 9.2 | 9.78 |
| 27 | 5.3 | 5.7 | 7.02 |
| 28 | 9.5 | 10 | 5.00 |
| 29 | 11 | 10 | 10.00 |
| 30 | 9.3 | 9.9 | 6.06 |
| 31 | 6.3 | 6 | 5.00 |
| 32 | 9.3 | 8.6 | 8.14 |
| 33 | 10 | 9.2 | 8.70 |

Table 2 (continued)**Table 2** (continued)

| SN | GA (weeks) | | Relative error (%) |
|----|------------|---------|--------------------|
| | Human | Machine | |
| 34 | 7.5 | 8.6 | 12.79 |
| 35 | 8.3 | 10 | 17.00 |
| 36 | 6.8 | 7.8 | 12.82 |
| 37 | 7.3 | 6.4 | 14.06 |
| 38 | 8.7 | 6.6 | 31.82 |
| 39 | 6.3 | 6.1 | 3.28 |
| 40 | 5 | 5.9 | 15.25 |
| 41 | 7.3 | 8 | 8.75 |
| 42 | 8.3 | 7.6 | 9.21 |
| 43 | 6.5 | 9.5 | 31.58 |
| 44 | 8 | 7.7 | 3.90 |
| 45 | 8 | 6.3 | 26.98 |
| 46 | 8.3 | 7.1 | 16.90 |
| 47 | 7 | 9 | 22.22 |
| 48 | 8.5 | 8.8 | 3.41 |
| 49 | 7.3 | 6.5 | 12.31 |
| 50 | 7.3 | 7.3 | 0.00 |
| 51 | 7.3 | 5.7 | 28.07 |
| 52 | 7.3 | 7.1 | 2.82 |
| 53 | 6.8 | 7.6 | 10.53 |
| 54 | 8 | 7.3 | 9.59 |
| 55 | 6.3 | 6.4 | 1.56 |
| 56 | 8.3 | 7 | 18.57 |
| 57 | 9.8 | 10.1 | 2.97 |
| 58 | 8.3 | 9.1 | 8.79 |
| 59 | 8.3 | 8.7 | 4.60 |
| 60 | 8.3 | 6.7 | 23.88 |
| 61 | 6.3 | 6.6 | 4.55 |
| 62 | 11 | 8.5 | 29.41 |
| 63 | 7 | 6.8 | 2.94 |
| 64 | 5.8 | 7.2 | 19.44 |
| 65 | 6.8 | 7.2 | 5.56 |
| 66 | 6 | 4.9 | 22.45 |

Table 2 (continued)

Table 2 (continued)

| SN | GA (weeks) | | Relative error (%) |
|----|------------|---------|--------------------|
| | Human | Machine | |
| 67 | 7 | 6.5 | 7.69 |
| 68 | 5 | 5.8 | 13.79 |
| 69 | 6.3 | 6.7 | 5.97 |
| 70 | 7.3 | 6.8 | 7.35 |
| 71 | 7.7 | 7.8 | 1.28 |
| 72 | 8 | 6.8 | 17.65 |
| 73 | 8 | 8.1 | 1.23 |
| 74 | 7.7 | 7.4 | 4.05 |
| 75 | 6.3 | 6.9 | 8.70 |
| 76 | 6.3 | 5 | 26.00 |
| 77 | 6.3 | 5.5 | 14.55 |
| 78 | 5 | 5.8 | 13.79 |
| 79 | 11 | 9.7 | 13.40 |
| 80 | 7.3 | 7.6 | 3.95 |
| 81 | 8.3 | 9 | 7.78 |
| 82 | 8.3 | 6.5 | 27.69 |
| 83 | 6.3 | 6.8 | 7.35 |
| 84 | 7 | 8 | 12.50 |
| 85 | 6 | 9.5 | 36.84 |
| 86 | 7.5 | 9.2 | 18.48 |
| 87 | 7.3 | 8.1 | 9.88 |
| 88 | 7.3 | 7.4 | 1.35 |
| 89 | 8.7 | 8.4 | 3.57 |
| 90 | 9.3 | 8.8 | 5.68 |
| 91 | 6.3 | 8.1 | 22.22 |
| 92 | 6.3 | 9.2 | 31.52 |
| 93 | 7.3 | 6.6 | 10.61 |
| 94 | 7.3 | 8.2 | 10.98 |
| 95 | 7.8 | 6.8 | 14.71 |
| 96 | 9.3 | 6.7 | 38.81 |
| 97 | 11.5 | 8.3 | 38.55 |
| 98 | 9 | 9.2 | 2.17 |
| 99 | 7.3 | 6.5 | 12.31 |

Table 2 (continued)

Table 2 (continued)

| SN | GA (weeks) | | Relative error (%) |
|-----|------------|---------|--------------------|
| | Human | Machine | |
| 100 | 12 | 10 | 20.00 |
| 101 | 8.3 | 6.3 | 31.75 |
| 102 | 8.3 | 6.9 | 20.29 |
| 103 | 8.3 | 8.5 | 2.35 |
| 104 | 6.5 | 8.3 | 21.69 |
| 105 | 8.7 | 7.1 | 22.54 |
| 106 | 8.3 | 8.8 | 5.68 |
| 107 | 9 | 7.9 | 13.92 |
| 108 | 7.5 | 6.3 | 19.05 |
| 109 | 8.3 | 8 | 3.75 |
| 110 | 6 | 6.7 | 10.45 |
| 111 | 7.3 | 7.5 | 2.67 |
| 112 | 5.5 | 6 | 8.33 |
| 113 | 6.5 | 7.8 | 16.67 |
| 114 | 7.3 | 6.8 | 7.35 |
| 115 | 7.3 | 5.5 | 32.73 |
| 116 | 8 | 6 | 33.33 |
| 117 | 6.3 | 5.5 | 14.55 |
| 118 | 9 | 7.2 | 25.00 |
| 119 | 5.5 | 7.4 | 25.68 |
| 120 | 8.3 | 6.8 | 22.06 |
| 121 | 8.3 | 6.8 | 22.06 |
| 122 | 7.5 | 6.2 | 20.97 |
| 123 | 8.3 | 8 | 3.75 |
| 124 | 8.3 | 8.4 | 1.19 |
| 125 | 4.8 | 6.3 | 23.81 |
| 126 | 4.8 | 5.5 | 12.73 |
| 127 | 8.3 | 6.7 | 23.88 |
| 128 | 7 | 6.1 | 14.75 |
| 129 | 8.5 | 9.6 | 11.46 |
| 130 | 6.5 | 8.4 | 22.62 |
| 131 | 7.3 | 8 | 8.75 |
| 132 | 7.3 | 7.8 | 6.41 |

Table 2 (continued)

Table 2 (continued)

| SN | GA (weeks) | | Relative error (%) |
|-----|------------|---------|--------------------|
| | Human | Machine | |
| 133 | 7.3 | 8.1 | 9.88 |
| 134 | 8.3 | 7 | 18.57 |
| 135 | 6.3 | 7.6 | 17.11 |
| 136 | 5.5 | 8 | 31.25 |
| 137 | 6.3 | 5.2 | 21.15 |
| 138 | 7.3 | 6.7 | 8.96 |
| 139 | 7.8 | 7.5 | 4.00 |
| 140 | 5 | 4.8 | 4.17 |
| 141 | 9 | 7.9 | 13.92 |
| 142 | 12 | 8.3 | 44.58 |
| 143 | 7.5 | 8.3 | 9.64 |
| 144 | 6.5 | 7.7 | 15.58 |
| 145 | 7.5 | 8.6 | 12.79 |
| 146 | 8 | 9.4 | 14.89 |
| 147 | 6.7 | 6.8 | 1.47 |
| 148 | 6.7 | 6.3 | 6.35 |

Table 2 (continued)

Table 2 (continued)

| SN | GA (weeks) | | Relative error (%) |
|-----|------------|---------|--------------------|
| | Human | Machine | |
| 149 | 7.5 | 8.4 | 10.71 |
| 150 | 8.5 | 9.7 | 12.37 |
| 151 | 8.3 | 7.8 | 6.41 |
| 152 | 7.5 | 9.1 | 17.58 |
| 153 | 9 | 8.7 | 3.45 |
| 154 | 7.3 | 7.8 | 6.41 |
| 155 | 7.3 | 7.3 | 0.00 |
| 156 | 7.3 | 8.3 | 12.05 |
| 157 | 8 | 6.2 | 29.03 |
| 158 | 7.3 | 8.9 | 17.98 |
| 159 | 6.3 | 6.7 | 5.97 |

The table shows the assessment of 1–159 video human-machine comparisons. “GA (weeks) Human” represents the gestational age assessed by the intermediate skills clinicians. “GA (weeks) Machine” represents the gestational age assessed by the new system. “Relative error (%)” represents the relative error assessed by the intermediate skills clinicians. “Absolute error (weeks)” represents the absolute error assessed by the intermediate skills clinicians. GA, gestational age; SN, serial number.

Table 3 One-sample statistics

| Error | n | Mean | Standard deviation | Standard error of the mean |
|-----------------------|-----|-------------|--------------------|----------------------------|
| Relative error (%) | 159 | 13.45% | 9.73% | 0.77% |
| Absolute error (week) | 159 | 1.0006 week | 0.76037 | 0.06030 |

Table 4 Independent sample *t*-test

| Error | Test value =0 | | | Mean difference (95% CI) |
|-----------------------|---------------|------|--------------------|--------------------------|
| | <i>t</i> | d.f. | P value (2-tailed) | |
| Relative error (%) | 17.431 | 158 | 0.000 | 13.45252 (11.93–14.98) |
| Absolute error (week) | 16.594 | 158 | 0.000 | 1.00063 (0.8815–1.1197) |

CI, confidence interval.

be implemented using a simple computer program, and the accuracy of the assessment using the automatic system was the same as the “gold standard”. The evaluation of GA in early pregnancy is a routine and frequent examination in hospitals (22), making it crucial to improve the efficiency

and accuracy of diagnosis.

The proposed system had certain limitations: some images extracted from the video were unclear and the boundary of the gestational sac area was blurred, making it difficult for the system to extract the contour. Further

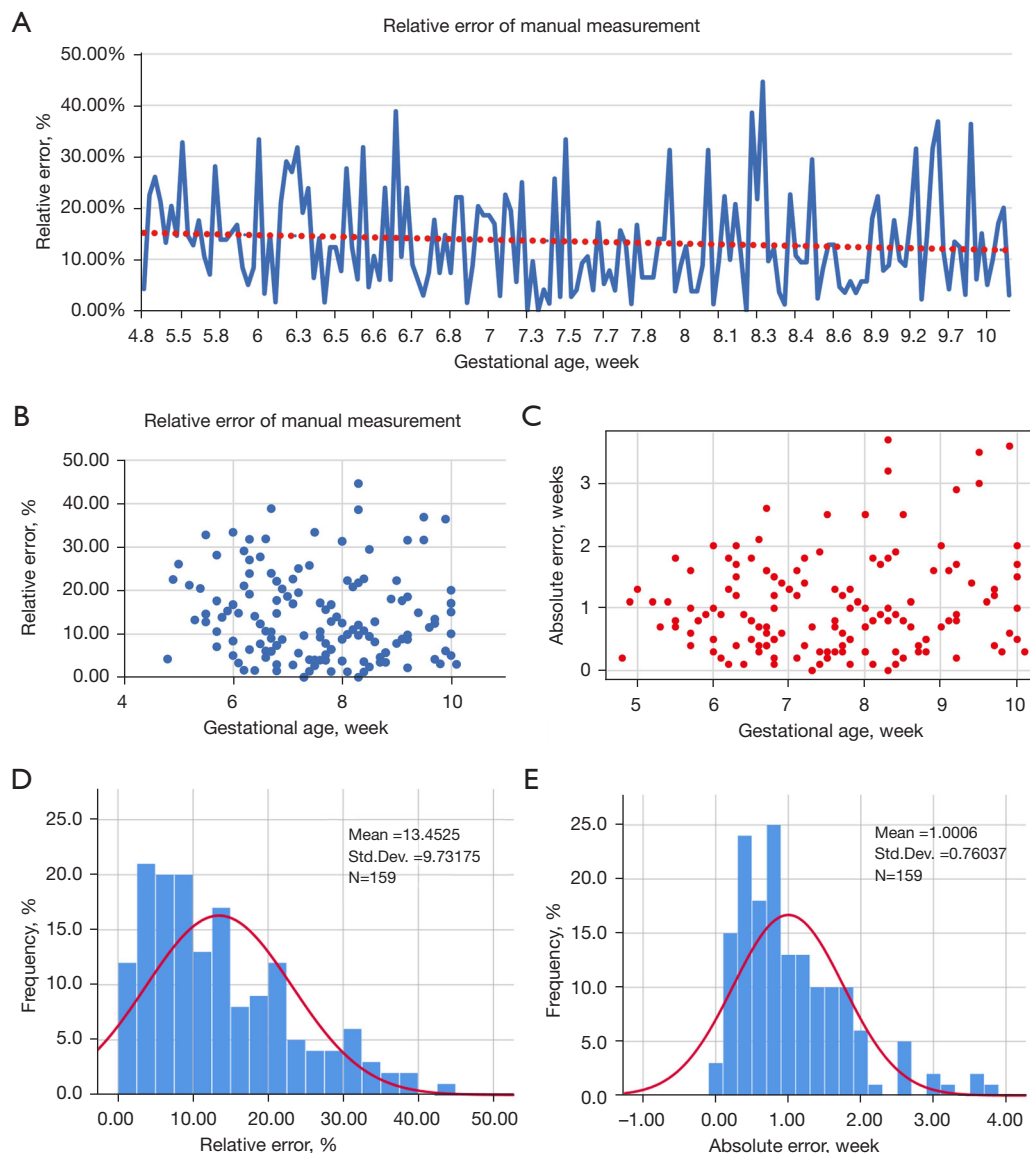


Figure 5 Analysis of clinician error and gestational age. (A) Relative error of manual assessment according to GA. The red dashed line means the trend line of relative error. (B,C) Scatter plots of relative (B) and absolute (C) errors. (D,E) Histograms showing the probability density distribution of relative (D) and absolute (E) errors. GA, gestational age.

research is needed in this area. Similar to any computer-aided diagnosis tool, it is recommended that clinicians perform “sanity checks” on a few special outputs (23,24). In addition, because a scale bar was used in the proposed method, the video had to be scanned at the same magnification. This study cannot eliminate errors caused by improperly oriented scans, and the new system cannot help with pathological diagnosis.

The GA was accurately predicted and evaluated by

estimating the maximum size of the gestational sac using 2D images. Although 3D ultrasound is better than 2D ultrasound in terms of the accuracy of fetal biometrics (25,26), 2D imaging exhibited solid clinical relevance, with the costs of B-ultrasound also lower.

In summary, this study has demonstrated the advantages of a new program that automatically captures the contours of the pregnancy sac. This program can also locate other lesion areas and measure their size to facilitate a “sanity

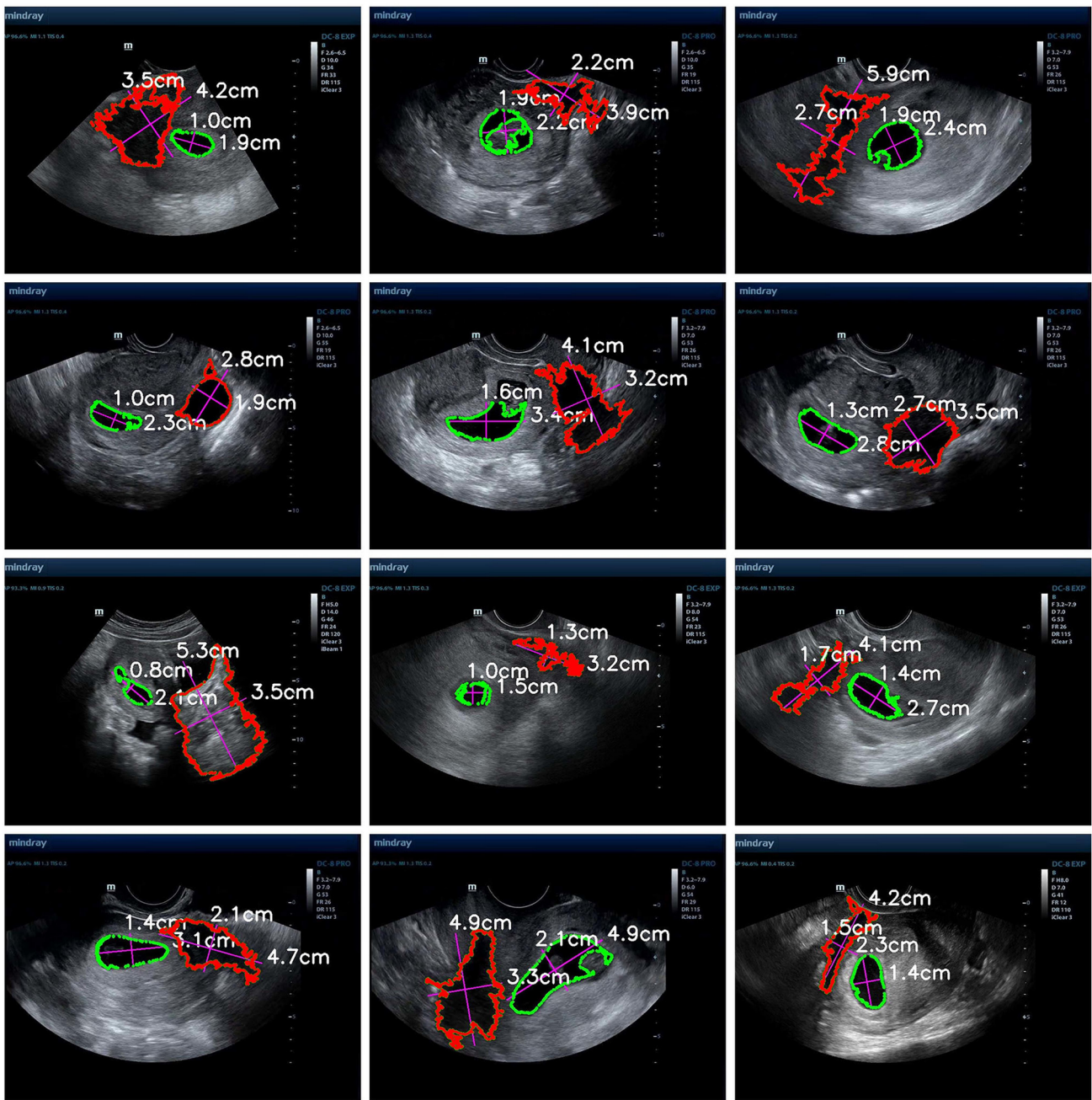


Figure 6 The “sanity checks” using the new system. Gestational sacs are outlined in green, whereas the areas of myomas, cysts, hematocel, scars, effusions, and other areas around the gestational sac are outlined in red.

check” for clinicians. This simple computer program integrates complex experimental procedures, it is easy to run on an ordinary personal computer, and it is convenient enough to be readily used for clinical diagnosis. The new biometric measurement method accurately measured the

maximum length and short diameter of the gestational sac, and then automatically calculated GA. Compared with the computer vision system, the clinicians’ assessment error of GA was 13.45%. Using the system, clinicians were able to obtain an accurate estimate of GA in early pregnancy

in approximately 30 s, compared with 15–20 min typically required for manual assessments. In addition, the system was able to evaluate the GA of B-ultrasound videos in batches, which was convenient for batch correction before giving the diagnosis report to patients.

The method proposed in this study could assist clinicians improve the efficiency and accuracy of GA calculation in early pregnancy.

Conclusions

Here, we automatically extracted the contour of the gestational sac using computer vision. Then, the new biometric measurement method automatically calculated the maximum length and short diameter of the pregnancy sac from multiple angles using the center of the pregnancy sac as a reference, with almost no error. Comparisons between the system and the intermediate skills clinician group showed that the efficiency and accuracy of the system were better than those of the intermediate skills clinicians. The system can also automatically estimate GA by measuring many B-ultrasound videos simultaneously. In addition, clinicians were able to complete the assessment of GA in 30 s, on average, making it a practical, repeatable, and reliable technique for clinical examinations in early pregnancy. This system would be useful for clinical application in routine and large-scale diagnoses, especially in basic rural hospitals where there is a shortage of experienced clinicians. This study also provided 159 video datasets and we have made them publicly available (<https://drive.google.com/drive/folders/1cP25UNROveiafvumT9vInQ2OQj3-xdug?usp=sharing>) so that more researchers can participate in improving and clinically verifying this method. In the future, we hope to expand and improve this computing framework through deep learning, artificial intelligence, and image segmentation (27–29).

Acknowledgments

The authors thank the editor and the anonymous reviewers for their constructive suggestions.

Funding: This work was supported by National Key R&D Program (No. 2019YFB1404803).

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-21-837/rc>

[com/article/view/10.21037/qims-21-837/rc](https://qims.amegroups.com/article/view/10.21037/qims-21-837/rc)

Conflicts of Interest: All authors completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-21-837/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The work was approved by the Ethics Committee of the Guangzhou Women and Children's Medical Centre (Scientific Research Ethics Committee permission No. GO-2016-017). Written informed consent was provided by all participants at the time of their initial hospital visit.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Parker JD, Schoendorf KC. Implications of cleaning gestational age data. *Paediatr Perinat Epidemiol* 2002;16:181–7.
2. Lynch CD, Zhang J. The research implications of the selection of a gestational age estimation method. *Paediatr Perinat Epidemiol* 2007;21 Suppl 2:86–96.
3. Butt K, Lim K; Diagnostic Imaging Committee. Determination of gestational age by ultrasound. *J Obstet Gynaecol Can* 2014;36:171–81.
4. Whitworth M, Bricker L, Mullan C. Ultrasound for fetal assessment in early pregnancy. *Cochrane Database Syst Rev* 2015;(7):CD007058.
5. Zhang L, Chen S, Chin CT, Wang T, Li S. Intelligent scanning: automated standard plane selection and biometric measurement of early gestational sac in routine ultrasound examination. *Med Phys* 2012;39:5015–27.
6. Chen H, Dou Q, Ni D, et al. editors. Automatic Fetal

- Ultrasound Standard Plane Detection Using Knowledge Transferred Recurrent Neural Networks. *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*. Cham: Springer International Publishing, 2015.
7. Ibrahim DA, Al-Assam H, Du H, et al. editors. Automatic segmentation and measurements of gestational sac using static B-mode ultrasound images. *ProcSPIE*; 2016:Vol. 9869 98690B-1.
 8. Kim HP, Lee SM, Kwon JY, Park Y, Kim KC, Seo JK. Automatic evaluation of fetal head biometry from ultrasound images using machine learning. *Physiol Meas* 2019;40:065009.
 9. Thevenot J, Lopez MB, Hadid A. A Survey on Computer Vision for Assistive Medical Diagnosis From Faces. *IEEE J Biomed Health Inform* 2018;22:1497-511.
 10. Burge M, Burger W. editors. Ear biometrics in computer vision. *Proceedings 15th International Conference on Pattern Recognition ICPR-2000*. 2000: IEEE.
 11. Yeung S, Rinaldo F, Jopling J, Liu B, Mehra R, Downing NL, Guo M, Bianconi GM, Alahi A, Lee J, Campbell B, Deru K, Beninati W, Fei-Fei L, Milstein A. A computer vision system for deep learning-based detection of patient mobilization activities in the ICU. *NPJ Digit Med* 2019;2:11.
 12. Khemasuwan D, Sorensen JS, Colt HG. Artificial intelligence in pulmonary medicine: computer vision, predictive model and COVID-19. *Eur Respir Rev* 2020;29:200181.
 13. Hellman LM, Kobayashi M, Fillisti L, Lavenhar M, Cromb E. Growth and development of the human fetus prior to the twentieth week of gestation. *Am J Obstet Gynecol* 1969;103:789-800.
 14. Suzuki S, be K. Topological structural analysis of digitized binary images by border following. *Comput Vis Graph Image Process* 1985;30:32-46.
 15. Bradski G, Kaehler A. *Learning OpenCV: Computer vision with the OpenCV library*: "O'Reilly Media, Inc."; 2008.
 16. Hinton P, McMurray I, Brownlow C. *SPSS explained*: Routledge, 2014.
 17. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab* 2012;10:486-9.
 18. Altman DG, Bland JM. Statistics notes: the normal distribution. *BMJ* 1995;310:298.
 19. Lin EP, Bhatt S, Dogra VS. Diagnostic clues to ectopic pregnancy. *Radiographics* 2008;28:1661-71.
 20. Shen Y, Luo J, Wang W. The Value of Prenatal Systematic Ultrasonic Examination of Fetal Structural Abnormality in Diagnosing Fetal Structural Abnormality. *J Med Imaging Health Inform* 2021;11:1623-32.
 21. Campbell S, Warsof SL, Little D, Cooper DJ. Routine ultrasound screening for the prediction of gestational age. *Obstet Gynecol* 1985;65:613-20.
 22. Goldenberg RL, Nathan RO, Swanson D, Saleem S, Mirza W, Esamai F, et al. Routine antenatal ultrasound in low- and middle-income countries: first look - a cluster randomised trial. *BJOG* 2018;125:1591-9.
 23. Milunsky A. Obstetrics, genetics, and litigation. *Acta Obstet Gynecol Scand* 2021;100:1097-105.
 24. Epstein H, Fleischer A. Sane Obstetrics. *JAMA* 1931;97:219-27.
 25. Roy-Lacroix ME, Moretti F, Ferraro ZM, Brosseau L, Clancy J, Fung-Kee-Fung K. A comparison of standard two-dimensional ultrasound to three-dimensional volume sonography for routine second-trimester fetal imaging. *J Perinatol* 2017;37:380-6.
 26. Ambroise Grandjean G, Berveiller P, Hossu G, Noble P, Chamagne M, Morel O. Prospective assessment of reproducibility of three-dimensional ultrasound for fetal biometry. *Diagn Interv Imaging* 2020;101:481-7.
 27. Zhu J, Zhang S, Yu R, Liu Z, Gao H, Yue B, Liu X, Zheng X, Gao M, Wei X. An efficient deep convolutional neural network model for visual localization and automatic diagnosis of thyroid nodules on ultrasound images. *Quant Imaging Med Surg* 2021;11:1368-80.
 28. Wan KW, Wong CH, Ip HF, Fan D, Yuen PL, Fong HY, Ying M. Evaluation of the performance of traditional machine learning algorithms, convolutional neural network and AutoML Vision in ultrasound breast lesions classification: a comparative study. *Quant Imaging Med Surg* 2021;11:1381-93.
 29. Zhang Y, Li H, Cao T, Chen R, Qiu H, Gu Y, Li P. Automatic 3D adaptive vessel segmentation based on linear relationship between intensity and complex-decorrelation in optical coherence tomography angiography. *Quant Imaging Med Surg* 2021;11:895-906.

Cite this article as: Pei Y, Gao W, E L, Dai C, Han J, Wang H, Liang H. Rapid and automatic assessment of early gestational age using computer vision and biometric measurements based on ultrasound video. *Quant Imaging Med Surg* 2022;12(4):2247-2260. doi: 10.21037/qims-21-837