Mathematical Biosciences and Engineering

Research article

Automatic measurement of fetal femur length in ultrasound images: a comparison of random forest regression model and SegNet

Fengcheng Zhu[1,†], Mengyuan Liu[1,†], Feifei Wang[2], Di Qiu[1,*], Ruiman Li[1] and Chenyang Dai[1]

[1] Department of Gynaecology and Obstetrics, the First Affiliated Hospital of Jinan University, Guangzhou, China

[2] Anesthesiology department, the First Affiliated Hospital of Jinan University, Guangzhou, China

† The authors contributed equally to this work.

* **Correspondence:** Email: qiudi1989@jnu.edu.cn.

Abstract: The aim of this work is the preliminary clinical validation and accuracy evaluation of our automatic algorithms in assessing progression fetal femur length (FL) in ultrasound images. To compare the random forest regression model with the SegNet model from the two aspects of accuracy and robustness. In this study, we proposed a traditional machine learning method to detect the endpoints of FL based on a random forest regression model. Deep learning methods based on SegNet were proposed for the automatic measurement method of FL, which utilized skeletonization processing and improvement of the full convolution network. Then the automatic measurement results of the two methods were evaluated quantitatively and qualitatively with the results marked by doctors. 436 ultrasonic fetal femur images were evaluated by the two methods above. Compared the results of the above three methods with doctor's manual annotations, the automatic measurement method of femur length based on the random forest regression model was 1.23 ± 4.66 mm and the method based on SegNet was 0.46 ± 2.82 mm. The indicator for evaluating distance was significantly lower than the previous literature. Measurement method based SegNet performed better in the case of femoral end adhesion, low contrast, and noise interference similar to the shape of the femur. The segNet-based method achieves promising performance compared with the random forest regression model, which can improve the examination accuracy and robustness of the measurement of fetal femur length in ultrasound images.

Keywords: ultrasonoscopy; femur length measurement; random forest; deep learning; prenatal diagnosis

1. Introduction

Medical Ultrasound imaging has been widely used in prenatal diagnosis due to its radiation-free, non-pioneering, real-time and low cost [1]. In the routine operation of pre-ultrasound examination, the physician usually needs to operate the built-in trackball for measurement. Such manual parameter measurement may have the following problems: the measurement of biological parameters is manually operated by the experienced sonographer, and repeated operation is very time-consuming, which increases the burden of ultrasound doctors [2,3]. Therefore, the obtained measurement results are greatly influenced by the subjectivity of the physician. Routine fetal biological parameters include Crown-rump length (CRL), biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC), and femur length (FL) [4]. The accurate measurement of fetal biological parameters is important for estimating the gestational age and weight of the fetus, as well as identifying the status of fetal development [5].

In recent years, most of the automatic measurement methods of the femur are the combination of image processing and traditional machine learning methods to extract image features and get the segmented image of the femur through traditional machine learning methods. With the continuous application of computer vision technology in medical images, deep learning is widely used in medical image analysis because of its good feature extraction and learning ability [6,7]. Considering the complex background of fetal femur ultrasound image, it is difficult to design features manually by traditional machine learning method. We also hope to use a deep learning method to measure femur automatically.

With the rapid development of computer vision and artificial intelligence technology, deep learning is widely used in the field of medical images. Sundaresan et al. [7] proposed a Fully Convolutional Neural Networks (FCN) network framework to assist in the screening of congenital heart disease by automatically analyzing fetal echocardiography, and successfully reduced the error rate of identification by 23.48%. Andermatt et al. [8] Used a three-dimensional recurrent neural network (RNN) to segment gray and white matter regions of the brain in MRI images; Chen et al. [6] proposed a transfer learning (transfer learning) framework based on a composite neural network to achieve the automatic detection of standard sections of fetal anatomy. Poudel et al. [9] combined a 2-dimensional U-Net structural and cyclic neural network (Gated recurrent unit) to describe 3-dimensional images and segment MRI cardiac images. Kroll et al. [10] proposed a CNN network based on the Hough voting mechanism, segmentation of brain MRI images showed that the learning-based segmentation method was robust and had the better generalization ability. Badrinarayanan et al. [11] first proposed the SegNet network in 2015, which modified the decoder of FCN, improved the progress of image segmentation, and reduced the memory occupancy during operation.

However, there are many challenges involved in quality assessment of the ultrasound images. The main difficulties of fetal femur measurement in ultrasound images are: 1) the inherent speckle noise, acoustic shadow, blurred edge imaging and other problems in ultrasound images, the difficulty of accurate segmentation of the femur [12,13], 2) the difference of the difference and the difference of the set parameters of the data collected by the doctor, resulting in large differences in the femoral figure, 3) clinically, the femoral endpoint is defined: the "middle point of U" shape at both ends of the femur, excluding the epiphysis [14]. However, most of the current research on automatic measurement methods of femur uses the most distal points of the segmented image as the femoral endpoint, which does not meet the clinical definition of femoral endpoint. Therefore, we hope to be able to design a method that can well solve the above difficulties and make the measurement results meet the clinical standards.

In this study, we implemented random forests method by a two-stage framework. Both stages were trained using a random forest regression model, and the only difference was that in the second stage of training, we would add automatic context features [15] to optimize the results of one stage. The predicted distance map of the femoral endpoint could be directly obtained by the above method. Next, the distance map could be post-processed utilizing the mean shift to obtain the final femoral endpoint coordinates, and then the femoral length was obtained by calculating the distance between the two points. However, traditional machine learning methods require manual design and selection of features, and it was necessary to verify whether the selection of features is appropriate through multiple experiments, which greatly improved the complexity of measurement methods. Therefore, we proposed the use of deep learning methods. First, we used fetal femoral ultrasound images and binary images of femoral contours marked by the sonographer as inputs and trained them with the SegNet model [11] to obtain the results after segmentation. Then the segmented results were processed, the midline of the segmented femur was obtained by the skeletonization method, and finally, the measurement results of the femur were obtained by calculating the distance between the two points. This paper compared the measurement results of the above two different methods to find the best method for automatic measurement of the femur.

2. Materials and methods

2.1. Automatic measurement method of femoral endpoint localization based on random forest regression model

In this part, we used random forest regression algorithm to train a two-stage regression model, which located the femoral endpoint directly without the need to segment the femoral image. The flowchart described is shown in Figure 1, and the details of the framework were presented in the following sections.

1) In the first stage, according to the gray level feature, location feature and gradient feature of the ultrasound image, the random forest regression algorithm was used to construct the mapping of these three features to the target structure.

2) The automatic context characteristics were added to the training of the two-stage regression model together with the one-stage regression model to achieve the optimization and improvement of the whole framework, and the final random forest regression model was obtained.

3) Finally, the distance graph with the clustering method was processed to obtain the final femoral endpoint coordinates.

Random forest was proposed in 1995 by Ho et al. [16] as an ensemble learning method, which is widely used in the field of image analysis and has achieved very good results. The basic idea is to construct multiple decision trees that are independent of each other during the training phase. When predicting input samples, random forest needs to integrate the predicted results of its various decision trees. The way regression and classification problems are integrated is slightly different. The classification problem, adopts a voting system, each decision tree votes to a category, and the category that obtains the most votes is the final result. In the regression problem, the prediction results obtained by each tree are real numbers, and the final prediction results are the average of the predicted results of each decision tree [17]. We hope to construct the regression model of the femoral endpoint by using the random forest to avoid the possible errors caused by the image segmentation and directly obtain
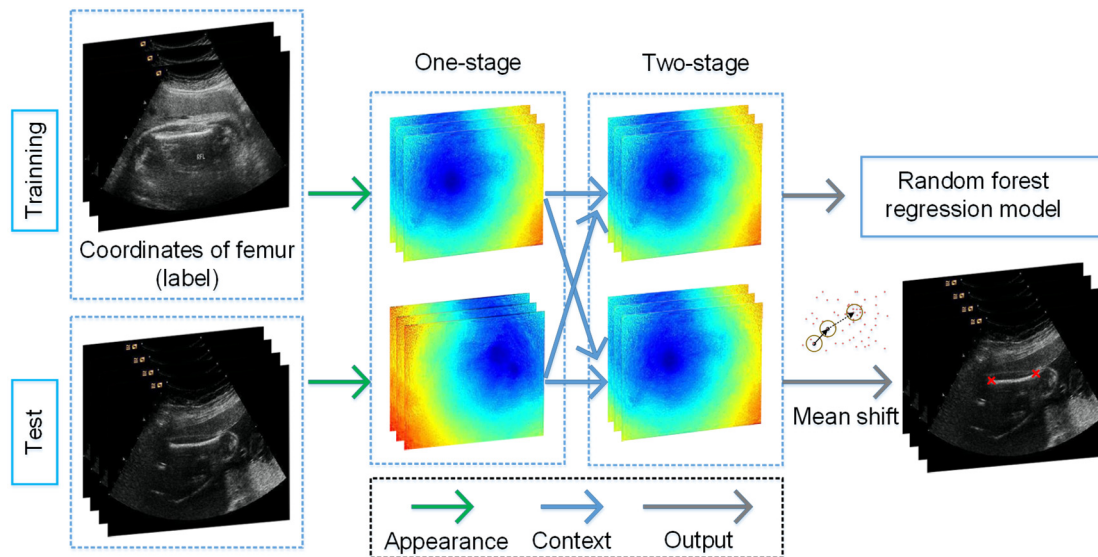
the localization results of the femoral endpoint.

Figure 1. Automatic positioning framework of femur based on random forest regression model.

2.1.1. Gaussian sampling

The features obtained by Gaussian sampling can get better results in the random forest regression task. We took the position of one femoral endpoint as the center of the Gaussian sampling, and the probability density function of the Gaussian distribution is given by:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \qquad (1)$$

where the μ represents the location parameter, and σ^2 = dag (60, 60) is variable. Then, the image was sampled according to the Gaussian function to obtain, the more sampling points, the closer to the target point, the more concentrated the sampling points, and conversely, the sparser the sampling points. In this way, the sampling point near the target point accounted for the proportion of all sampling points, which could describe the characteristics of the target point better.

2.1.2. Feature extraction

To construct a regression model of the femur endpoints, we extracted three types of features for the training of the random forest regression model.
- Gray features, which is composed of the central gray level feature of a pixel i in the image, and the gray level feature within its surrounding neighborhood. 16 rays evenly distributed around the periphery, centered on sampling point b [18]. On each ray, set the sampling point, according to the way of Gaussian distribution. By collecting the sampling points with Gaussian distribution on these rays, the gray level characteristics of the sampling point b are obtained. To make the feature more robust to the speckle noise points in the ultrasound image, the final feature value of each sampling point is the mean of the gray levels in the local window of 5 × 5 centered on the current

sampling point. In this method, the dimension of grayscale features is 91-dimensional. For a certain pixel i in the image, its grayscale feature can be expressed as $\{gray_p(i), 1 \leq p \leq 91\}$.

- Location features, in the ultrasound image, the two endpoints of the femur are distributed on the left and right sides of the image, respectively [19]. Therefore, we divide the coordinate values (x (i), y (i)) of a pixel i in the image by the length and width of the current image, respectively, to obtain the normalized coordinate value (lx (i), ly (i)), which is added to the feature vector. This feature dimension is 2-dimensional.

- Gradient features, the contrast information about the image is extracted, and similarly, for the robustness of the feature expression. So, again we computed the average of the gradient in the local window centered on the sampling point 5×5 as the feature value of this point. At the very end, the gradient feature is denoted as $\{grad_p(i), 1 \leq p \leq 90\}$, and the gradient feature dimension is 90 dimensions.

In summary, a total of 183-dimensional features was obtained as the input for the one-stage random forest regression model.

2.1.3. One-stage regressor training

In one-stage random forest regression model training, we calculated the feature vector and the corresponding annotation to obtain the training data matrix. This data matrix was used as input and put into the random forest for training to obtain the regression model. After obtaining the input features, we denoted the left and right endpoints of the femur by $d^l(i)$ and $d^{lr}(i)$, respectively, and constructed a nonlinear map between image pixel i and these two target structures. This nonlinear feature map was obtained by random forest regression. We set the tree of the decision tree in the random forest regression model to be 30 and the minimum number of samples in the leaf nodes in each tree to be 5, and we used a parallel way to train. When the random forest learning was completed, the random forest regressor would output the distance between image pixel i and two target structures, and finally got a predicted distance map. The regression model for each target structure needed to be trained twice, and finally, we needed to perform four random forest training sessions in the one-stage training.

2.1.4. Two-stage regressor training

However, there was a certain positional structural link between the two endpoints of the femur. The information describing the connection was added to the feature space in this method, and the original image features are enhanced using automatic context features to achieve the effect of optimizing the one-stage random forest regressor. We obtained the two- stage random forest regression model by iterative training and tested the test data with this model. As shown in Figure 2.

The automatic context model proposed by Tu [20] and others had a good performance in tasks such as image segmentation and image recognition [21]. The core idea of this algorithm was the cascade superposition of a series of training models. The predicted results outputed by the upper-level regressor were fused with the gray level feature, location feature and gradient feature to obtain the more effective features compared with the upper-level feature, which were input into the current regressor, and then the prediction results are refined, which was continuously iterated in this way until the optimal prediction results were obtained.
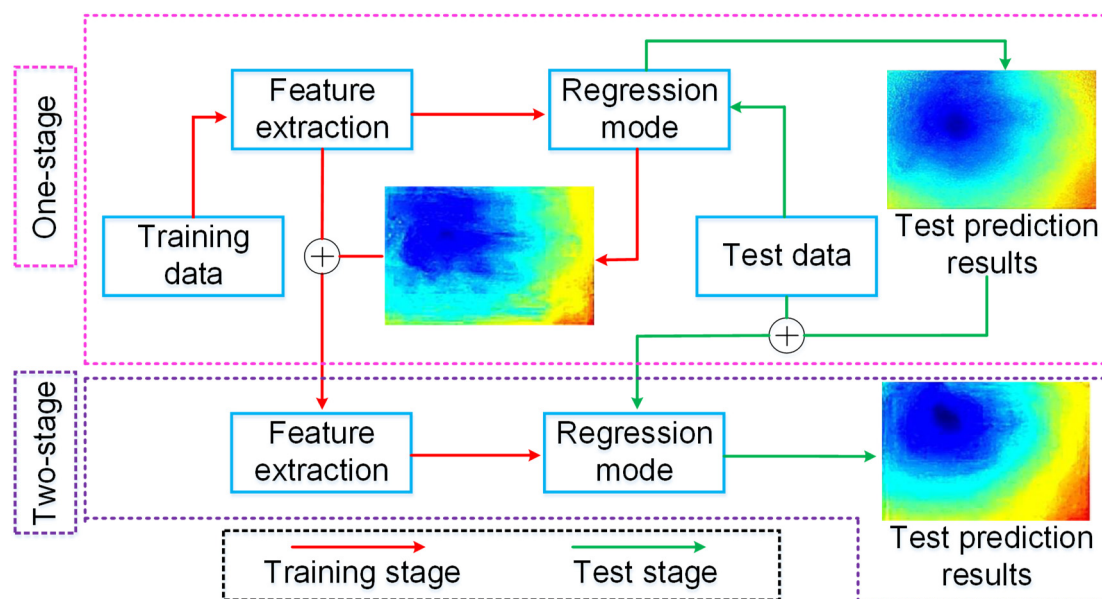
Figure 2. Automated context optimization framework.

In one-stage random forest regression model, the training of the two endpoints of the femur was completely independent, and the interconnection between these structures was not taken into account during the training. Through the understanding of the context characteristics, we believed that the contextual relationship between the two target structures will be helpful in the training of the regressor, such as the two endpoints must be left and right distributed. In this method, the one-stage regressor was trained to obtain the distance features $r^L(i)$ and $r^R(i)$ of pixel i in the image, which was used to extend the original feature vector $\mathbf{x} = \{\{gray_p\}, lx, ly; \{gray_{p'}\}, 1 \leq p \leq 91, 1 \leq p' \leq 90\}$. After contextual feature enhancement, the feature vector can be expressed as $\mathbf{x}_{context} = \{\{gray_p\}, lx, ly; \{gray_{p'}\}, \{r_p^L(i)\}, \{r_p^R(i)\}; 1 \leq p \leq 91, 1 \leq p' \leq 90\}$, and it is 365-dimensional.

The enhanced feature vector was re-input into the random forest for regression training and the number of iterations is set to 1, and many studies [22] have shown that setting the number of iterations to 1 is sufficient. In the femoral endpoint localization, we first predicted the initial distance map of each target structure in the ultrasound image with a step-by-step forest regressor. Contextual information was extracted for each target structure. After combining the features and contextual features of the original image, we predicted the final distance map for each target structure with a second-order random forest regressor.

2.1.5. Femoral endpoint positioning

After the above process, we got the prediction graph of the test image. For the femur endpoint P, we could get it by searching the position with the smallest distance in the corresponding distance graph. However, such a method was easily disturbed by noise. Therefore, we used the technique of mean shift [19] to determine the location of the target point. Mean shift was first proposed by Fukunaga et al. [23] in 1975 and originally meant as a mean vector of excursions [24]. With the development of

this theory, the current mean shift represented an iterative step. Specifically, first, we calculated the mean offset of a random point, then moved the point to its mean offset, used this as the new starting point, and continued moving until certain conditions were met [24], as shown in Figure 1.

2.2. Automatic femur measurement method based on SegNet

The femur automatic measurement method frame based on SegNet [11] is shown in Figure 3. The framework of this method is based on the deep learning method, which takes the SegNet network trained in the end-to-end mode as the core to directly realize the segmentation of the whole ultrasound image. Compared with the traditional machine learning algorithm, SegNet can learn effective features from the data by itself, without the need for manual feature design and selection. First, we cut the original picture to remove the irrelevant ultrasound equipment information; then, we enhanced the ultrasound data to solve the problem that the network is difficult to train due to the small number of ultrasound; next, during the test, we could obtain the segmented image of the femur by simply entering the picture to be tested into SegNet; finally, we extracted the femoral bone endpoint information and calculated the femur length through the skeletonised post-processing method.
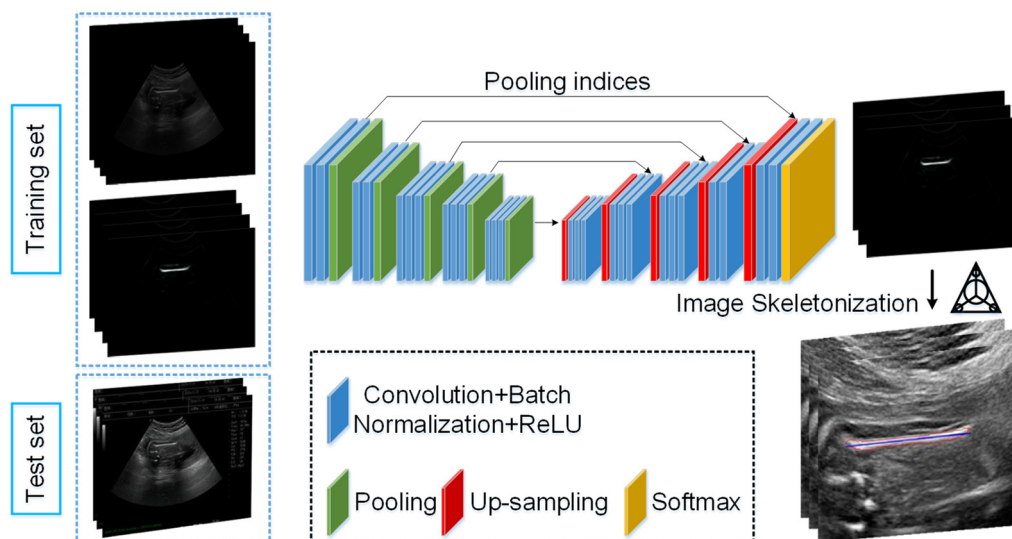

Figure 3. Automatic measurement framework of femur based on SegNet-model.

2.2.1. SegNet

SegNet was an image semantic segmentation deep network proposed by Badrinarayanan et al. [11] to solve autopilot or intelligent robot. As shown in Table 1, SegNet was similar to FCN [25] and consisted of an encoder and a symmetry-structured decoder, where the encoder used the first 13-layer convolution network of VGG-16, each encoder layer corresponds to a decoder layer, and the final decoder output was fed into the softmax classifier on the last layer to get the maximum probability of each pixel in all categories.

Table 1. The architecture of SegNet.

Network layer name	Convolution kernel size	Step
C1	$3 \times 3 \times 64$	-
C2	$3 \times 3 \times 64$	-
Max-pooling	2×2	2
C3	$3 \times 3 \times 128$	-
C4	$3 \times 3 \times 128$	-
Max-pooling	2×2	2
C5	$3 \times 3 \times 256$	-
C6	$3 \times 3 \times 256$	-
C7	$3 \times 3 \times 256$	-
Max-pooling	2×2	2
C8	$3 \times 3 \times 512$	-
C9	$3 \times 3 \times 512$	-
C10	$3 \times 3 \times 512$	-
Max-pooling	2×2	2
C11	$3 \times 3 \times 512$	-
C12	$3 \times 3 \times 512$	-
C13	$3 \times 3 \times 512$	-
Max-pooling	2×2	2

The encoders of SegNet (convolution layer, pooling layer, activation function) are based on the translation without deformation and rely only on relative spatial coordinates. For an image pixel sitting mark x (i, j) in a specific layer X, the corresponding mapping of its output feature map in the image pixel y (i, j) corresponding to the next layer Y is:

$$Y_{ij} = f_{ks}(\{X_{si+\delta i, sj+\delta j}\} 0 \leq \delta i, \delta j \leq k) \tag{2}$$

where the k is the convolution kernel size; s is the step size; $f_{ks}(\cdot)$ is the activation function, and ReLU is used in this paper. ReLU is an improvement of the traditional activation function sigmoid, which can well solve the problem of gradient disappearance. The output of ReLU is a = max (0, z). So we defined the loss function of SegNet as:

$$\mathrm{l}(\mathrm{x}; \theta) = \sum_{i,j} l'(x_{i,j}; \theta) \tag{3}$$

where the x is the training set data and θ is the parameter of SegNet. Each convolution layer is composed of a convolution plus a batch normalization layer (Batch normalisation, Bn) and activation layer (ReLU) composition in SegNet. The main role of the Bn layer is to speed up learning, and when training and testing, the process of action of the bn layer in the encoder and decoder can be summarized as:

1) During training: (a) Forward propagation: the Bn layer standardizes the convolved eigenvalues, but the Bn layer only stores the mean and variance of the input eigenvalues, and the value when the eigenvalues are output to the next convolution layer is still the eigenvalue of the previous convolution output. (b) Backward propagation: According to the mean and variance of the eigenvalues stored in the Bn layer, each convolution layer is chained with the ReLU layer to obtain the gradient so as to

calculate the current learning rate.

2) During the test: the Bn layer obtains the mean and variance of all the data in the training set, and then calculates its output according to the uncompiled estimation of the whole in the training set.

The up-sampling process of SegNet is to map the value of the feature map into the feature map of the row through the previously saved maximum poolable coordinates, and set the other positions to zero. Compared with the FCN network, SegNet is superior to FCN in terms of segmentation accuracy and memory occupancy [26].

2.2.2. Skeletonization analysis

We could get the segmented image of fetal ultrasound femur directly through SegNet network. By calculating the intersection of the skeleton curve and the femur contour of the segmented image, the final endpoint of the femur was located. It is important that the femoral endpoints are positioned accurately to meet clinical criteria in the femoral measurements. Since the gold standard for femoral endpoints is the midpoint of the "U" shaped region at both ends of the femur, it is necessary to determine the central axis of the femur, and skeletonization will be used in this study to find the central axis of the femoral region. The method of describing an image with a skeleton was first proposed by Blum et al. [27] to define the skeleton with the concept of the largest disk. If A was not a subset of any inner tangent circle in, then A was called the largest disk in the image, and at this point, we defined skeleton C as the set of the centers of all largest disks in the image. According to the definition of skeletonization based on the maximum disk method, the distance from the skeleton point to the disk and the image tangent point were the same, which ensured the axial characteristics of the skeleton well [28]. We labeled the two points furthest from the intersection point of the skeletonization curve and the femoral contour as the femoral endpoint. As shown in Figure 3, once the femoral endpoint is confirmed, we can calculate the distance between the two points to obtain the femoral length value.

2.3. Experimental setup

In this paper, we collected a total of 435 fetal femoral ultrasound images with a size of 1200×900. This method was directly validated experimentally using 435 raw data. First, we preprocessed the ultrasound image and removed the surrounding equipment information, and the size of the processed ultrasound image was 768×576. Due to the limited number of collected ultrasound images, we enhanced these ultrasound images by taking vertical images (considering that horizontal images may affect the position information of the two ends of the femur), randomly taking four different angles (two positive and two negative) within the range of $\pm 30°$ for rotation, and finally obtaining 2610 raw data. We divided the 2610 image data into two parts: the training set and the test set, of which 2300. Since this method directly locates the femoral endpoint and does not need to segment the image, we only evaluate the position of the femoral endpoint and the position marked by the doctor and the femoral length in the evaluation of this method.

Random forest regression: to evaluate the results of this method, we need to label the femoral endpoint and extract the coordinates of the femoral endpoint in the ultrasound image. In this method, we asked a well-trained sonographer majoring in medical image processing to mark the points at both ends of the femur, and submitted the mark results to another sonographer for verification to ensure the accuracy of the bid winning of the training data.

SegNet: segmented images are needed to provide as a label for training, and when evaluating the results of this method, we also need segmented images to evaluate our segmentation results. The segmented image for each data annotation is like a binary image with a foreground region labeled as 1 and background region labeled as 0. We asked a sonographer to label the data to ensure the accuracy of labeling in the training data. The image distributions of the training and test sets are consistent with the automatic femur measurement method based on random forest regression model.

In addition, we asked an ultrasonographer to annotate the femoral contour and femoral endpoint data to ensure the accuracy of the annotation in the evaluation of the method results. This experiment was all run under Ubuntu14.04 with 2.20 GHz Intel ® Xeon ® E52650 CPU and 256G memory.

3. Experiments and results

3.1. Detection results

Figure 4 shows the error curves of random forests in both stages. The abscissa is the number of trees in the random forest decision tree, the ordinate is the training error of the random forest, the blue curve represents the error curve of the one-stage random forest, and the red curve represents the error curve of the two-stage random forest. It can be seen that after combining the automatic contextual characteristics, the error of the random forest regression model can converge faster, while the training error is smaller (close to 1500). By comparing the error curves of the random forest regression model in the two-stages, it can be concluded that the automatic context feature improves the model in one-stage to a certain extent, which proves the effectiveness of the automatic context feature. We can get two-stage random forest regression corresponding to two target structures, which can be used in the two-stage data test together with one-stage random forest regression. The two-stage random forest regression model was obtained by iterative training. Table 2 shows the quantitative evaluation of the automatic measurement results of the femur in the random forest regression model. It can be seen that the ratio of the end point of the femur and the length of the femur in the two stages are significantly lower than that in the first stage ($p < 0.05$).
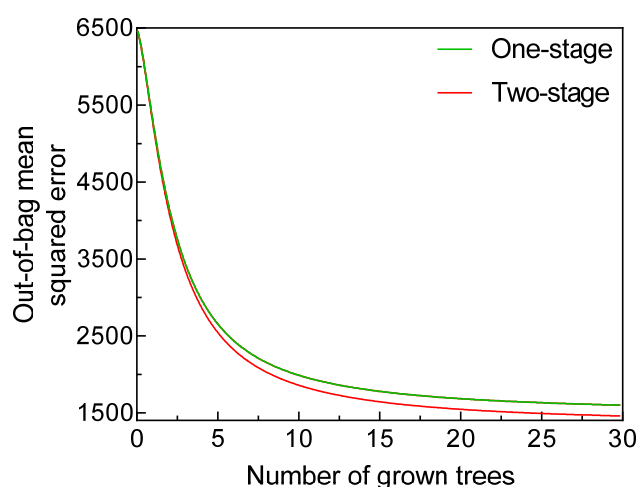

Figure 4. Training error of random forest in one-stage and two-stage.

Table 2. Quantitative evaluation of automatic measurement of femur based on random forest regression model.

	Comparison of left endpoint of femur (mm)	Comparison of right endpoint of femur (mm)	Comparison of length of femur (mm)
One-stage	3.93 ± 4.03	3.45 ± 5.62	2.01 ± 5.81
Two-stage	2.87 ± 3.38	2.52 ± 4.76	1.23 ± 4.66

Figure 5. Qualitative comparison of the different methods. Red: random forest regression model. Blue: SegNet. Green: physician-annotated.

In this study, we used 2300 fetal femur ultrasound images for training and used 310 fetal femur ultrasound images as the test set. We compared the measurement results with the physician-annotated results. Figure 5 shows the comparison of the measurement results of the two methods, where red is the automatic measurement method of femur based on random forest regression model, blue is the automatic measurement method of femur based on SegNet, and green is the physician-annotated results (gold standard). As can be seen, the SegNet-based automatic femur measurement method still performs well in cases of highlighted noise of similar femoral structures, low image contrast and sticking of the femur ends with the noise. The accuracy and robustness of the SegNet based method for automatic measurement of femurs were demonstrated.

We used the Bland-Altman index [29] to evaluate the consistency between this method and the results labeled by the physician. The degree of consistency between the two methods is that the dashed line representing the mean difference is about close to the solid line representing the mean difference of 0 [29].

95% Upper Limit = Mean of difference + 1.96 × Standard deviation of the mean difference

Figure 6(a) Bland-Altman plot of the based-method. From the figure, it can be seen that the mean difference between the results of this method and those labeled by the doctor is 1.23 mm, and the upper and lower limits of 95% are 10.36 and -7.90 mm. From Eq (7), the standard deviation of the mean difference between the results of this method and those labeled by the doctor is 4.66 mm. The results show that this method has good consistency with the results labeled by the doctor.

Figure 6(b) shows the comparison of the agreement between the automatic measurement method of femur based on SegNet and the results labeled by the physician. From the figure, it can be seen that the mean difference between this method and the physician labeled results is 0.46 mm, and the upper

and lower limits of 95% are 4.71 and -3.8 mm, respectively. The closer the upper and lower limits of 95% are to the mean difference, the better the consistency between the two methods. The results show that this method has very good consistency with the physician labeled results.
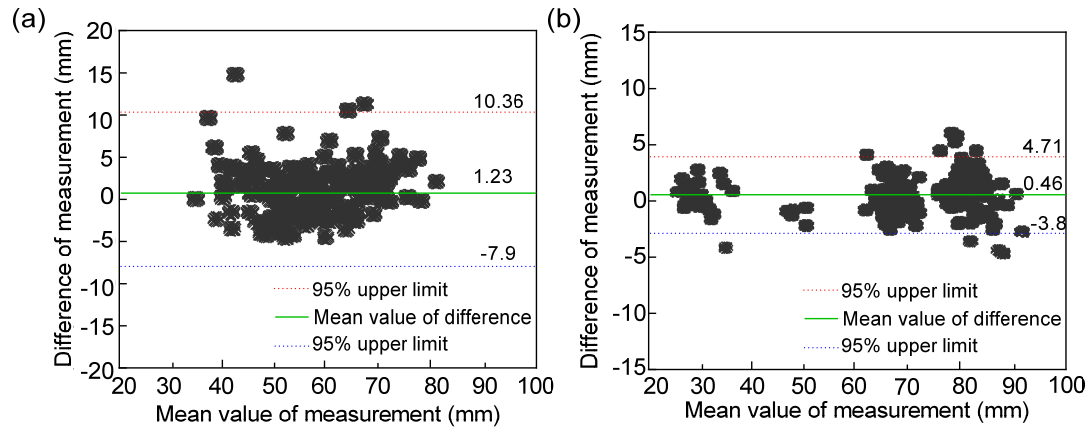

Figure 6. Bland-Altman plot of measurements based on random forest regression model (a) and SegNet (b).

Table 3 lists the femur length measurement and that for two methods. The mean value of difference of SegNet (0.46 mm) was almost 3 times smaller than that of random forest regression measurement (1.23 mm), demonstrating the efficiency and stability of the automatic method.

Table 3. Comparison of the femur length between random forest regression method and our SegNet.

Method	Comparison of femur length (mm)
Random forest regression	1.23 ± 4.66
SegNet	0.46 ± 2.82

We used area-based and distance-based evaluation metrics to verify the accuracy of the segmentation results of the two methods' femur auto dynamometry methods. Area-based evaluation indicators are usually used to compare the differences between automatic segmentation and manual segmentation regions, including Precision, Specificity, Sensitivity and Dice [30]. The distance-based evaluation indicators compare the differences between automatic segmentation results and physician-annotated contours, including MSD (maximum symmetric contour distance), ASD (average symmetric distance) and RMSD (root mean contour distance).

Comparison of maximum entropy segmentation method in Wang's study [31] and femur segmentation result based on SegNet model. From table 4, it can be seen that the femur segmentation results based on SegNet model are significantly better than the maximum entropy segmentation method, and the accuracy of the femur automatic measurement method based on SegNet is proved from the perspective of quantitative evaluation of segmentation effect.

Table 4. Quantitative comparison of automatic measurement methods of femur based on SegNet model.

Evaluating indicator (segmentation image)		Literature [31]	SegNet model
Area	Precision /%	74.66 ± 9.32	86.06 ± 8.73
	Specificity /%	83.72 ± 13.11	92.11 ± 7.03
	Sensitivity /%	99. 80 ± 0.22	99.86 ± 0.13
	Dice /%	80.27 ± 8.44	92.2 ± 6.71
Distance	MSD /mm	6.02 ± 7.29	1.97 ± 1.03
	ASD /mm	1.04 ± 1.29	0.05 ± 0.12
	RMSD /mm	1.77 ± 2.41	0.19 ± 0.21

4. Discussion

Prenatal ultrasonography is an irreplaceable method to observe the growth and development of the fetus. Biological parameters have an important influence on the accuracy of judging fetal growth and development. At present, the measurement of biological parameters is manually operated by experienced ultrasound doctors. Therefore, the measurement results obtained are very affected by the subjectivity of doctors and rely heavily on the experience of ultrasound doctors. So, the automatic measurement of fetal biological parameters is of great significance in prenatal ultrasound. First of all, it can accurately measure the length of the femur, which is more efficient and economical than the manual operation of doctors. Second, the measurement results are more objective and avoid the possible errors of the physicians with different experiences.

In this paper, the automatic measurement method of fetal femur was selected, which was based on the automatic measurement algorithm of the random forest regression model and SegNet. We combined machine learning with image processing to achieve automatic measurement of fetal ultrasound femurs. For the automatic measurement method of femur, this method directly located the femoral endpoint for the first time, and directly located the femoral endpoint through the random forest regression model, which avoided the secondary error that may be caused by segmentation to a certain extent. At the same time, the automatic context method was used to improve the efficiency of the model. Subsequently, we combined the automatic measurement algorithm with the SegNet framework to obtain the segmentation results of the femur; then, the segmentation results were processed to obtain the final femur length; finally, we compared the results of the two methods with the physician annotation, and the results showed that the automatic measurement algorithm of the femur based on the SegNet model was more in good agreement with the physician annotation, meeting the clinical requirements.

It can be found that the results based on SegNet model are in better agreement with the labeling results of doctors than the random forest regression model. Compared with other automatic measurement methods [32], this method works better on pictures with low resolution, it also has a more accurate positioning. However, due to the random forest regression model method has the common problem of machine learning, that is, the characteristics need to be manually designed and selected, so that the complexity of the measurement method is greatly improved. Therefore, we hope to design a deep learning method that can automatically learn the features of the image, reduce the trouble caused by the manual design features, and reduce the complexity of the automatic measurement method for the femur.

However, the automatic measurement method based on SegNet performed better in the results, due to the small data set of images (only 435 cases), the rest was enhanced from these 435 original data and were not tested in the larger data set. Therefore, we would like to be able to test on a larger number of datasets to get a more persuasive result. Whether it is random forest regression-based automatic measurement algorithm or SegNet-based automatic measurement algorithm, although the overall results are better, but there is further room for improvement in the details of coordinate position. The location of the femoral endpoints differed somewhat from the results annotated by the physicians, especially based on the random forest regression model approach. Therefore, we hope to be able to locate the endpoints of the femur more precisely in future algorithm development. No matter the algorithm based on random forest regression or the algorithm based on SegNet, although it performs well in the overall results, there is still room for further improvement in the details of coordinate position. There is a certain difference between the position of the femur endpoint and the results marked by doctors, especially based on the random forest regression model. Therefore, we hope that in the future algorithm development, we can locate the endpoint of femur more accurately.

5. Conclusions

In this paper, we compared two automatic techniques for quality assessment of fetal femur in ultrasound images. SegNet network was used for fetal ultrasound image segmentation for the first time, which effectively overcomes the interference of similar bright structures in fetal femoral ultrasound images and improves the accuracy of localization and robustness to image quality. However, the speed of automatic measurement algorithm based on the random forest regression model has been greatly improved, it still fails to meet the requirements of real-time. In this regard, we hope to achieve the purpose by further optimization procedure or GPU. Systematic verification only uses the labeled results of sonographers as the gold standard for automatic measurement. However, manual measurement by sonographers can be subject to error, so the mean value of the results independently annotated by clinicians is required as the gold standard for automatic measurement.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. W. J. Cong, J. Yang, D. N. Ai, H. Song, G. Chen, X. H. Liang, et al., Global patch matching (GPM) for freehand 3D ultrasound reconstruction, *BioMed. Eng. OnLine*, **16** (2017), 216–214.
2. D. T. Avalokita, T. Rismonita, A. Handayani, A. W. Setiawan, Automatic fetal head circumference measurement in 2D ultrasound images based on optimized fast ellipse fitting, in *Tencon 2020-2020 IEEE Region 10 Conference (Tencon)*, IEEE, 2020.
3. M. van Tulder, A. Malmivaara, B. Koes, Repetitive strain injury, *Lancet*, **369** (2007), 1815–1822.
4. R. Gaillard, E. A. P. Steegers, J. C. de Jongste, A. Hofman, V. W. V. Jaddoe, Tracking of fetal growth characteristics during different trimesters and the risks of adverse birth outcomes, *Int. J. Epidemiol.*, **43** (2014), 1140–1153.

5. S. Lou, K. Carstensen, I. Vogel, L. Hvidman, C. P. Nielsen, M. Lanther, et al., Receiving a prenatal diagnosis of Down syndrome by phone: a qualitative study of the experiences of pregnant couples, *BMJ Open*, **9** (2019), e026825.

6. R. Qu, G. Xu, C. Ding, W. Jia, M. Sun, Standard plane identification in fetal brain ultrasound scans using a differential convolutional neural network, *IEEE Access*, **8** (2020), 83821–83830.

7. V. Sundaresan, C. P. Bridge, C. Ioannou, J. A. Noble, Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks, in *2017 IEEE 14th International Symposium on Biomedical Imaging*, IEEE, 2020.

8. S. Andermatt, S. Pezold, P. Cattin, Multi-dimensional gated recurrent units for the segmentation of biomedical 3d-data, in *Deep Learning and Data Labeling for Medical Applications*, Springer, 2016.

9. R. P. K. Poudel, P. Lamata, G. Montana, Recurrent fully convolutional neural networks for multi-slice mri cardiac segmentation, in *Reconstruction, Segmentation, and Analysis of Medical Images*, Springer, Cham, (2016), 83–94.

10. C. Kroll, F. Milletari, N. Navab, S. A. Ahmadi, Coupling convolutional neural networks and hough voting for robust segmentation of ultrasound volumes, in *German Conference on Pattern Recognition*, Springer, Cham, (2016), 439–450.

11. V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for scene segmentation, *IEEE Trans. Pattern Anal.*, **39** (2017), 2481–2495.

12. S. Dahdouh, E. D. Angelini, G. Grange, I. Bloch, Segmentation of embryonic and fetal 3D ultrasound images based on pixel intensity distributions and shape priors, *Med. Image Anal.*, **24** (2015), 255–268.

13. J. A. Noble, D. Boukerroui, Ultrasound image segmentation: A survey, *IEEE Trans. Med. Imaging*, **25** (2006), 987–1010.

14. L. J. Salomon, Z. Alfirevic, V. Berghella, C. Bilardo, E. Hernandez-Andrade, S. L. Johnsen, et al., Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan, *Ultrasound Obst. Gyn.*, **37** (2017), 116–126.

15. P. Hu, F. Wu, J. Peng, D. Kong, Automatic 3d liver segmentation based on deep learning and globally optimized surface evolution, *Phys. Med. Biol.*, **61** (2016), 8676–8676.

16. T. K. Ho, Random decision forests, in *Proceedings of 3rd international conference on document analysis and recognition*, IEEE, (1995), 278–282.

17. J. Gall, V. Lempitsky, Class-specific Hough forests for object detection, in *IEEE Conference on Computer Vision & Pattern Recognition*, IEEE, 2009.

18. A. Biswas, S. Dasgupta, S. Das, A. Abraham, A synergy of differential evolution and bacterial foraging optimization for global optimization, *Neural Netw. World*, **17** (2007), 607–626.

19. D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, *IEEE Trans. Pattern Anal.*, **24** (2002), 603–619.

20. Z. W. Tu, X. G. Chen, A. L. Yuille, S. C. Zhu, Image parsing: unifying segmentation, detection, and recognition, *Int. J. Comput. Vision*, **63** (2005), 113–140.

21. Y. Gao, D. Shen, Context-aware anatomical landmark detection: application to deformable model initialization in prostate CT images, in *Machine Learning in Medical Imaging*, Springer, Cham, (2014), 165–173,

22. Y. Lu, H. P. Chan, J. Wei, L. M. Hadjiiski, Selective-diffusion regularization for enhancement of microcalcifications in digital breast tomosynthesis reconstruction, *Med. Phys.*, **37** (2010), 6003–6014.

23. K. Fukunaga, L. Hostetler, The estimation of the gradient of a density function, with applications in pattern recognition, *IEEE Trans. Inf. Theory*, **21** (1975), 32–40.

24. W. Liang, X. Xie, J. Wang, Y. Zhang, J. Hu, A SIFT-based mean shift algorithm for moving vehicle tracking, in *2014 IEEE Intelligent Vehicles Symposium Proceedings*, IEEE, 2014.

25. G. Crichton, S. Pyysalo, B. Chiu, A. Korhonen, A neural network multi-task learning approach to biomedical named entity recognition, *BMC Bioinf.*, **18** (2017), 318–332.

26. V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal.*, **39** (2017), 2481–2495.

27. S. Robert, Models for the perception of speech and visual form: Weiant Wathen-Dunn, *J. Commun. Disord.*, **1** (1968), 342–343.

28. A. McKnight, D. Si, K. Al Nasr, A. Chernikov, N. Chrisochoides, J. He, Estimating loop length from CryoEM images at medium resolutions, *BMC Struct. Biol.*, **13** (2013).

29. J. M. Bland, D. G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, *Lancet*, **1** (1986), 307–310.

30. S. Rueda, S. Fathima, C. L. Knight, M. Yaqub, A. T. Papageorghiou, B. Rahmatullah, et al., Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge, *IEEE Trans. Med. Imaging*, **33** (2014), 797–813.

31. C. W. Wang, Automatic entropy-based femur segmentation and fast length measurement for fetal ultrasound images, in *2014 International Conference on Advanced Robotics and Intelligent Systems*, IEEE, 2014.

32. P. Mukherjee, G. Swamy, M. Gupta, U. Patil, K. B. Krishnan, Automatic detection and measurement of femur length from fetal ultrasonography, in *SPIE: Medical Imaging 2010: Ultrasonic Imaging, Tomography, and Therapy*, SPIE, 2010.