Journal of Imaging Informatics in Medicine (2024) 37:247–267
https://doi.org/10.1007/s10278-023-00908-8

Fetal Ultrasound Segmentation and Measurements Using Appearance and Shape Prior Based Density Regression with Deep CNN and Robust Ellipse Fitting

Gaurav Dubey[1] · Somya Srivastava[2] · Anant Kumar Jayswal[3] · Mala Saraswat[4] · Pooja Singh[5] · Minakshi Memoria[6]

Received: 3 May 2023 / Revised: 5 September 2023 / Accepted: 6 September 2023 / Published online: 12 January 2024

Abstract

Accurately segmenting the structure of the fetal head (FH) and performing biometry measurements, including head circumference (HC) estimation, stands as a vital requirement for addressing abnormal fetal growth during pregnancy under the expertise of experienced radiologists using ultrasound (US) images. However, accurate segmentation and measurement is a challenging task due to image artifact, incomplete ellipse fitting, and fluctuations due to FH dimensions over different trimesters. Also, it is highly time-consuming due to the absence of specialized features, which leads to low segmentation accuracy. To address these challenging tasks, we propose an automatic density regression approach to incorporate appearance and shape priors into the deep learning-based network model (DR-ASPnet) with robust ellipse fitting using fetal US images. Initially, we employed multiple pre-processing steps to remove unwanted distortions, variable fluctuations, and a clear view of significant features from the US images. Then some form of augmentation operation is applied to increase the diversity of the dataset. Next, we proposed the hierarchical density regression deep convolutional neural network (HDR-DCNN) model, which involves three network models to determine the complex location of FH for accurate segmentation during the training and testing processes. Then, we used post-processing operations using contrast enhancement filtering with a morphological operation model to smooth the region and remove unnecessary artifacts from the segmentation results. After post-processing, we applied the smoothed segmented result to the robust ellipse fitting-based least square (REFLS) method for HC estimation. Experimental results of the DR-ASPnet model obtain 98.86% dice similarity coefficient (DSC) as segmentation accuracy, and it also obtains 1.67 mm absolute distance (AD) as measurement accuracy compared to other state-of-the-art methods. Finally, we achieved a 0.99 correlation coefficient (CC) in estimating the measured and predicted HC values on the HC18 dataset.

Keywords Fetal head (FH) · Head circumference (HC) estimation · Ultrasonic (US) imaging · Image segmentation · Measurement · Density regression · Deep learning networks

Introduction

Ultrasonic (US) imaging remains a significant screening tool for monitoring and detecting abnormalities in fetal growth during pregnancy. It has the advantage of recording trimester scans using video clips for future reference [1]. US images clearly identify the scanning reports, such as the pregnant woman's name, identification number, date of the US examination, birth date, and examination site (private scanning center or hospital) [2, 3]. The collection of the first trimester (0–13 weeks) provides a week-by-week examination of embryonic and fetal development, exposing all the intricate details of fetal growth. A unique peak of fetal development inside the womb was generated to reveal all

✉ Somya Srivastava
somyasrivastava215@gmail.com

1 Department of Computer Science, KIET Group of Institutions, Delhi-NCR, Ghaziabad, U.P, India

2 ABES Engineering College, Ghaziabad, U.P, India

3 Amity University, Noida, Uttar Pradesh, India

4 Department of Computer Science, Bennett University, Greater Noida, India

5 Shiv Nadar University, Greater Noida, Uttar Pradesh, India

6 CSE Department, UIT, Uttaranchal University, Dehradun, Uttarakhand, India

the intricate details of fetal growth in the second trimester (14 to 27 weeks). Next, the third trimester (28 to 42 weeks) refers to the full growth of fetal which is its full turn and preparation for birth [4, 5]. Based on these findings, the sonographer will take trimester measurements to estimate fetal growth [6].

Nowadays, there are lots of biometric measurements to take during the US examination. In the report of the fetal size examination, some measurements are included: abdominal circumference (AC), biparietal diameter (BPD), femur length (FL), and head circumference (HC) [7, 8]. Estimated delivery date (EDD) can be determined by combining BPD and HC measurements. In previous studies, HC predicted the date of confinement better than BPD [9]. HC is one of the significant biometric measurements for estimating gestational age (GA) and monitoring the growth of the fetus in US images. Also, it measures the outside of the head skull in the form of an ellipse; fewer than three standard deviations from the mean could be a sign of microcephaly [10, 11]. In clinical routine, an automatic measurement of fetal biometrics is performed on US images. The majority of the current computer-aided detection (CAD) systems use fetal HC measurement techniques under the premise that the FH contour is roughly elliptical [12]. A crucial step in assessing HC is identifying the FH contour, but due to flaws in fetal US imaging, such as artifacts, attenuation, and speckle noise, it can be challenging to automatically segment the FH from US images. These issues make it difficult to segment fetal US images due to the blurred head, fetal skull interruption by US artifacts, and interference from structures that have a texture similar to that of the FH, like the uterine wall interface and a lot of amniotic fluid [13–15].

Several existing techniques have been employed for FH segmentation from US images in the domain of computer vision applications, including multi-level thresholding, the Jaya algorithm, Chan-Vese [16], morphological operators, and texture maps [17]. However, these methods might not be able to detect FH in low-contrast US images. Other machine learning (ML) techniques, such as the random forest (RF) classifier [18], were applied for the detection of incomplete ellipses. However, it consumes more time to learn all the features and has a high misclassification rate. Recently, deep learning (DL) technology has been a popular research area to combine biomedical images with artificial intelligence in the domain of computer vision applications [19, 20]. Inside of DL, convolutional neural networks (CNNs) were the most popular network model for segmentation, especially in medical images [21]. Some of the representative CNN architectures include fully CNN (FCNs), referred to as Mini Link-Net [22], U-Net [23], and 3DV-Net [24]. The regression-based work regression CNNs, such as VGGNet and VGG16 [25, 26], are also used as a backbone network to extract patch characteristics for segmentation from the US images. Despite the advances in technology, automatic segmentation remains a challenging task due to the varying pixel sizes of fetal US images throughout all trimesters of pregnancy. Furthermore, there is a need for improvement in the effectiveness of current deep learning-based algorithms for extracting FH measurements.

Motivated by the above facts and existing studies, this paper proposes a DR-ASPnet model with robust ellipse fitting to automatically segment and measure FH from US images. The main idea is to determine the complex location of FH based on appearance and shape priors for segmentation and the elliptical parameters from FH profiles for HC estimation. Three developed frameworks jointly learn appearance and shape features during training and testing. Then the REFLS approach with robust weight functions is used to increase the accuracy of the ellipse parameter solution. Hence, our segmentation model is greatly enhanced in terms of robustness and determines the best fit for HC estimation with high accuracy.

Main Contributions

We present a DR-ASPnet model with robust ellipse fitting for fetal US images. This model is able to determine the complex location of FH for segmentation and determine the elliptical parameters from FH profiles for HC estimation.
We present a HDR-DCNN network model for automatic FH segmentation. Our network model involves three developed frameworks:

- Firstly, the US images are explicitly fed into the complex density regression network to extract the appearance of black pixel information and determine the complex location of FH.
- Secondly, the depth map of the US image is explicitly fed into the hierarchical density regression network to sufficiently extract shape information. Then retrain both features to locate the complex FH.
- Thirdly, both appearance features and shape features are used to train the deep CNN (DCNN) network to achieve the final segmentation result.

We present a measurement strategy using a REFLS approach to determine the ellipse parameters of FH and avoid incomplete ellipse problems for HC estimation.
Our proposed model validation results will show better than the other methods on the publicly available HC18 US dataset.

The rest of the paper is organized as follows: the "Related Works" section reviewed several recent methods in the domain of fetal US images. The "Proposed Methodology" section presents the proposed methodology for FH

segmentation and FH measurement, with a network framework diagram elaborated in detail. Simulation results and analysis are presented in the "Experiments and Evaluation" section. Lastly, the overall conclusion is given in the "Conclusion and Future Scope" section.

Related Works

Several recent works carried out in the domain of US images for FH detection, measurement, and segmentation are discussed as follows: To avoid manual measurement of fetal HC, Li et al. [27] introduced an automatic ML method using RF and fast ellipse fitting (ElliFit) in US images. This initial knowledge is fed as input to the RF classifier. Then the shape of the HC ellipse is fitted for measurement using ElliFit. Whereas another author, Van den Heuvel et al. [28] offered an automated assessment of fetal HC measurement for all trimesters using ML techniques in two-dimensional (2D) US images. Thus, the method achieves superior results for all trimesters when simulated on a large dataset. However, the testing process was performed by manual annotation measurement. Another author, Skeika et al. [21], developed a DL based V-Net combination (VNet-c) architecture to segment the fetal skull and measure the HC in 2D US images. The combination strategy employed 3D V-Net as the base network of the fully convolutional neural network (V-Net FCN) to learn the features with segmentation accuracy. The architecture used eight steps to mitigate overfitting problems. Fiorentino et al. [29] introduced regression-based CNNs (R-CNNs) to solve the edge delineation problem for HC measurement in US fetal images. Two edge-localization structures were trained with different factors for edge-localization tasks. Another author, Zeng et al. [30], presented a V-Net architecture-based deep supervised attention-gated (DAG V-Net) model in 2D US images. After obtaining fetal HC measurements through elliptical fitting, it is observed that the DAG V-Net model exhibits quicker convergence compared to the U-Net and V-Net models. Amini et al. [31] presented an automatic extraction method to determine the FH parameters using DL in multi-scale US images. The fetal extraction step used a deep link network (Deep-LinkNet) model to segment the FH for three trimesters, reducing the parameters using a loss function. To increase the generalization ability of multi-source US data, Zhou et al. [32] introduced a Fourier domain adaptation (FDA)-based learning network to improve the segmentation performance of FH. The FDA approach was used to learn the target domain by avoiding complex adversarial training. Thus, the model provides adaptive migration parameters for the task of generalization.

From the above-reviewed articles, the existing methods have some limitations, such as incomplete ellipses, variable location of the head, and fluctuations due to FH dimensions over different trimesters in the available datasets. Also, variations in different pixel sizes lead to challenges in segmentation performance during training. It is time-consuming due to the absence of specialized knowledge about different image conditions. Further, false localization, limited training features, and insufficient elliptical parameters lead to lower segmentation and measurement accuracy for fetal US images.

Proposed Methodology

This paper presents a DR-ASPnet model with a robust ellipse fitting method for FH segmentation and HC estimation in US images. Firstly, we highlight a clear view of significant features through image pre-processing and augmentation steps for training and testing a set of fetal US images. For automatic segmentation, the HDR-DCNN model integrates complex density regression and hierarchical density regression with deep convolutional network design. Then eliminate small irregular structures and smooth the segmented output using post-processing. Lastly, we applied the REFLS approach to compute the HC of each US image and determine the best fitting position of FH according to the elliptical parameters. The methodological framework of the DR-ASPnet model is shown in Fig. 1.

Data Description

The HC18 Grand Challenge [28] is a publicly available dataset that involves the collection of fetal US images validated for the purpose of training and evaluating machine learning models for fetal biometric measurements. The dataset was created by the Department of Obstetrics database at the Radboud University Medical Center in Nijmegen, Netherlands. A total of 1334 two-dimensional (2D) US images were obtained from 551 pregnant women at different stages of pregnancy. These images represent the standard plane and can be utilized for fetal HC assessment. The ground truth images in this dataset are medical US images of the head that have been manually labeled by human experts to provide accurate measurements of the head circumference. Using the random division approach, we randomly split the dataset into two subsets, usually in an 80:20 ratio for the training and testing sets, respectively, for three types of trimesters. The training set involves 999 images that split 165, 693, and 141 images for the first, second, and third trimesters, respectively, while the testing set involves 335 images that split 55, 223, and 47 images for the first, second, and third trimesters, respectively. Each 2D ultrasound image has a resolution of 840×500 pixels, and the size of each pixel varies from 0.052 to 0.326 mm. As the benchmark for automatic image

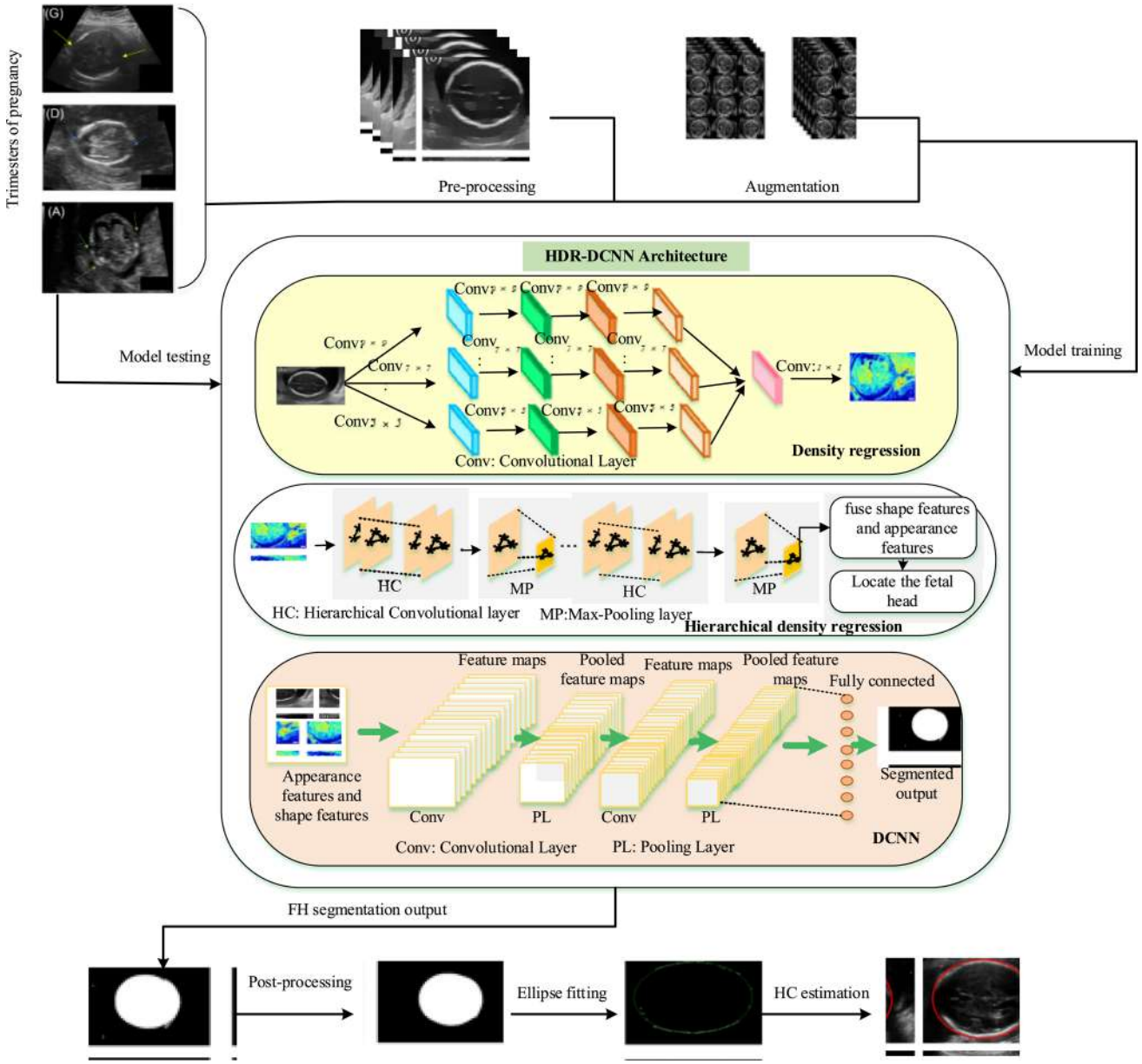


Fig. 1 Methodological framework for FH segmentation and HC measuremen

segmentation, the training set also includes HC images that have been manually label-ed by expert radiologists. A segmentation gold standard is, however, missing from the testing set. The findings from the testing set need to be evaluated against HC18 in order to determine the DR-ASPnet model performed in the HC18 Challenge (https://hc18.grand-challenge.org/). When dealing with multiple highly correlated US fetal images, we need to take into account the dependency between the images during development and testing. To address the correlation between the images, we perform data pre-processing techniques, as described in the "Phase 1: Image Pre-Processing and Augmentation" section. This helps reduce the influence of intensity variations across the images and makes them more comparable. Also, to introduce more diversity into the dataset and reduce overfitting, we apply data augmentation techniques specifically designed for correlated data, as described in the "Phase 1: Image Pre-Processing and Augmentation" section. Some of the sample images in the training set for three trimesters are illustrated in Fig. 2.

The DR-ASPnet methodological framework for FH segmentation and HC estimation involves three main phases:

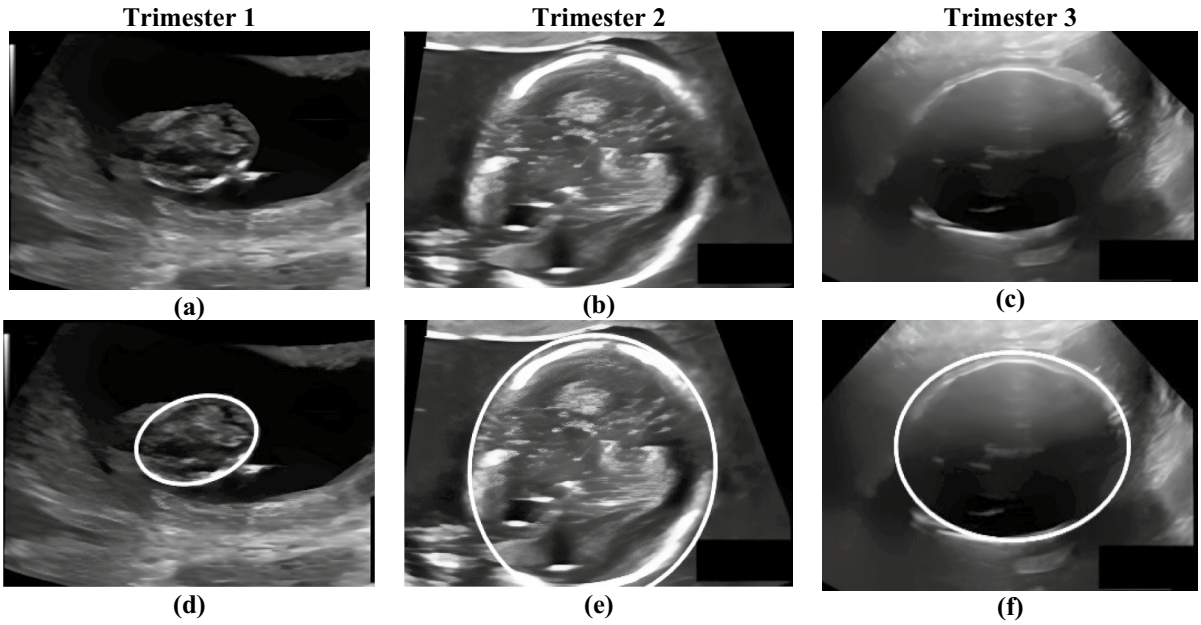


Fig. 2 **a–f** Representation of US images and ground truth images from the HC18 dataset (where **a–c** represent the three 2D US images and **d–f** its respective ground truth images).

Phase 1: Image Pre-Processing and Augmentation

Initially, the obtained fetal US images from the database are categorized into three trimesters of scan images. In our work, we labeled trimester scans (three trimesters of pregnancy) manually from the entire dataset for testing and training conditions. In these stages, manual annotations are provided by the medical experts for diagnostic treatment. Medical experts involved in the interpretation and analysis of fetal US images may include individuals at various stages of their medical careers, such as medical students, residents, fellows, and board-certified physicians specializing in relevant disciplines, like expert radiologists. In the context of medical experts and qualifications related to fetal US images, here are some common grades or qualifications: Residents undergo further training in a specific medical specialty, such as expert radiologists. During their residency, they work in hospitals or clinics under the supervision of attending physicians and gain hands-on experience in various aspects of patient care, including interpreting and analyzing medical images like fetal US images. A fellow is a physician who may undergo a fellowship in fetal medicine or maternal–fetal medicine. These fellowships focus on advanced training in the diagnosis and management of complex fetal conditions, including the interpretation of fetal US images. Fellows work closely with experienced specialists in the field and gain expertise in handling high-risk pregnancies and fetal abnormalities.

Normally, captured images have unique problems, such as blurred images over different trimesters and variations in different pixel sizes. To solve these issues, the pre-processing phase is used to remove unwanted distortions, variable fluctuations, and a clear view of significant features from the US images using three processing steps.

(i) **Normalization:** Generally, the intensity range of US images is 0 to 255. To shift and rescale the US image pixel-wise intensity, we employed normalization process through changing pixel intensity values within the range of 0 and 1 [33]. This process has the advantage of solving image pixel variations. The formula for expressing the normalization value (Nz) in the image as:

$$Nz = \frac{I_o - I_{MIN}}{I_{MAX} - I_{MIN}} \tag{1}$$

where I_o is the original image value and I_{MIN} and I_{MAX} represent the minimum and maximum image values, respectively.

(ii) **Resizing:** The image size is 800 × 540 pixels. To simplify the burden of proposed model, the input images are uniformly re-sized into 512 × 512 pixel resolution for training process. After resizing, we applied a threshold technique to maintain the pixel values in an image to a specific range, usually between 0 and 255. In thresholding, we define a threshold value that separates the darker pixel values from the lighter pixel values in the image. We then rescale the pixel values by mapping the darker

pixel values to 0 and the lighter pixel values to 255, with all other pixel values scaled proportionally in between. This ensures that the pixel values in the resized image are within the range of 0 and 255 [34].

(iii) **Binary encoding:** This step is often used to encode categorical variables into numerical values by encoding categories as integers and transforming them into binary code (i.e., encode the values [0 and 1], where 0 denotes black pixels and 1 denotes white pixels). During the training process, we represent the class in black pixels as (1, 0) and white pixels as (0, 1) [35]. This step has the advantage of reducing the curse of dimensionality.

In the case of fetal US images, data augmentation is used to create additional training samples that simulate variations in image acquisition conditions and fetal positions [36]. The augmentation operation involves center flipping, horizontal flipping, rotation, brightness, and contrast adjustment in the US images. In this case, center flipping involves flipping the image around its center point. It can simulate different fetal positions along the sagittal plane. Horizontal flipping involves flipping the image along the horizontal axis. This can simulate different fetal positions along the transverse plane. Brightness and contrast adjustments are used for different lighting conditions.

- ***Flipping***: Center flip = −10° to +10°; horizontal flip = −20° to +20°
- ***Brightness***: −20 to +20%
- ***Contrast***: ±10% to ± 20%
- ***Rotation***: Randomly rotated up to 10° to 15°

Phase 2: Automatic Segmentation

This step applied the HDR-DCNN model as an extractor process that takes appearance and shape prior features of fetal US images as input and represents a binary mask of segmented FH as output. There are three network models in the proposed HDR-DCNN model:

(i) **Appearance-Based Complex Density Regression** Initially, the appearance of black pixel features is extracted from the US images using complex density regression. Next, we validate US images to determine the complex location of FH with prediction error. In this, the complex density regression network model involves three blocks: convolutional layers, dense layers, and the density regression layer. A series of convolutional blocks are used to extract appearance features at different scales. The output of the convolutional blocks is flattened and fed into two dense layers to extract high-level features that capture complex relationships between appearance features. Then the output of the dense layers is fed into a complex density regression layer that estimates the distribution of the fetal contour in polar coordinates. This layer uses complex-valued weights to capture both magnitude and phase information, which is employed to estimate the final density map. The design of the complex density regression network allows for the unified integration of feature extraction into the density regression process, thereby eliminating the need for additional feature extraction methods. The framework of complex density regression is shown in Fig. 3.

(i) Appearance Based Complex Density Regression:

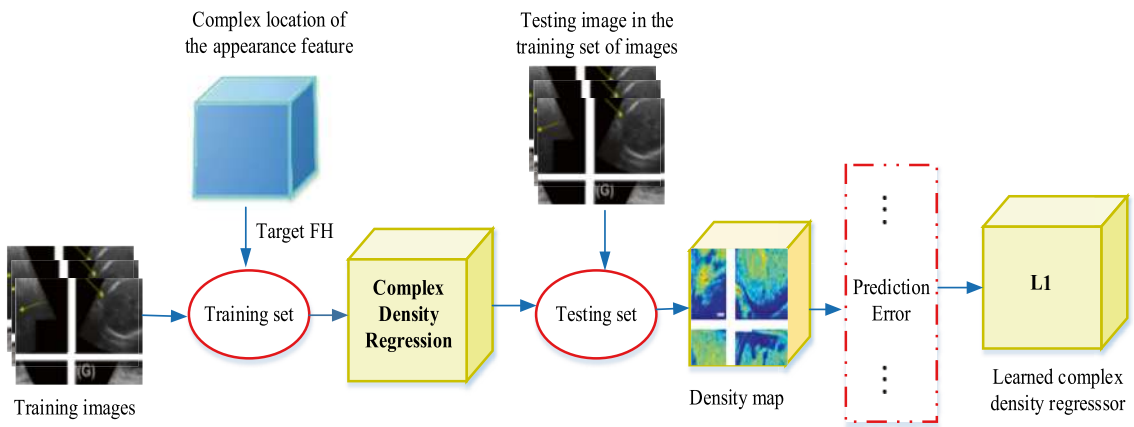


Fig. 3 Framework of complex density regression (L1)

Table 1 Terms and definitions related to fetal anatomy for first trimester (weeks 0–13)

Terms	Definitions
Gestational sac	A fluid-filled structure in the uterus that surrounds and supports the developing embryo
Yolk sac	A small sac attached to the embryo that provides nourishment in the early stages of pregnancy
Embryo	The developing organism from conception until the end of the 10th week of pregnancy
Fetal pole	The early visualization of the embryo with distinct features, such as a developing head and body
Nuchal translucency	A measurement of fluid accumulation at the back of the baby's neck, used for screening purposes
Placenta	An organ that develops during pregnancy and provides oxygen and nutrients to the fetus
Fetal heartbeat	The rhythmic contractions of the fetal heart, usually detected using Doppler ultrasound

The given $H \in ¥^{U\times V}$ be the testing image and the training complex density regression function is denoted as $f(H;\Theta)$, where U × V represents the number of rows and columns in the image, respectively. The corresponding function is used to estimate the density map for H, which is represented as $\hat{D}= (f(H;\Theta))$. Thus, the complex location of the appearance prior in the testing images is estimated $(\hat{C}_l)$ in the below equation:

$$\hat{C}_l= \sum_{i=1}^{U}\sum_{j=1}^{V}\hat{D}_{i,j} = \sum_{i=1}^{U}\sum_{j=1}^{V}\left[\hat{D}= L1(f(H;\Theta))\right]_{i,j} \quad (2)$$

where the expression $\left[\hat{D}= L1f(H;\Theta)\right]_{i,j}$ denotes the estimated density map of the pixel at location (i,j) in the image H.

The prediction error for each location in the appearance prior is estimated by:

$$Q_m = \overline{q_m} + \hat{q}_m \quad (3)$$

where:

$$\overline{q_m} = \frac{1}{K}\sum\nolimits_{k=1}^{K} q_m^k$$

and

$$\hat{q}_m= \sqrt{\frac{1}{K}\sum\nolimits_{k=1}^{K}\left(q_m^k - \overline{q_m}\right)^2}$$

where K represents the collection of US fetal anatomy at different trimesters, as well as annotated images that highlight important landmarks and structures. Some common terms and definitions related to fetal anatomy for three different trimesters in the context of US imaging are outlined in the Tables 1, 2, and 3.

(ii) **Appearance and Shape-Based Hierarchical Density Regression** Hierarchical density regression (HDR) framework is used to capture shape features, such as the shape and size of the FH features, in a hierarchical manner and to generate a shape representation of the whole model based on density map information. The hierarchical structure means a sequentially connected density regression model. For preparing the training samples, there is a corresponding density map for each image. In this, each density regression is trained on appearance features from US images and shape features from density maps. Then retrain the complex head location of fetal with the new shape features. Finally, the learned density regression of each iteration is applied to hierarchically locate the head during the testing process. The framework of HDR is shown in Fig. 4.

The hierarchical representation for density regression is obtained by fusing the target shape representations from various skull feature scales in the task-specific prediction module. Following this, a multilayer perceptron (MLP) is used to leverage the hierarchical representation for the density regression task. The blocks used in the HDR task for

Table 2 Terms and definitions related to fetal anatomy for second trimester (weeks 14–27)

Terms	Definitions
Amniotic fluid	The fluid that surrounds the fetus in the amniotic sac. It helps protect and cushion the fetus
Fetal movements	The movements felt by the mother as the baby starts to develop coordinated muscle activity
Fetal organs	The major organs, such as the heart, lungs, brain, kidneys, and liver, that continue to develop
Sex determination	The ability to determine the biological sex of the fetus through ultrasound imaging
Fetal growth	Measurements of the fetus, including head circumference, abdominal circumference, and femur length
Fetal position	The orientation of the fetus within the uterus, typically described in terms of head or breech presentation

Table 3 Terms and definitions related to fetal anatomy for third trimester (weeks 28–42)

Terms	Definitions
Braxton Hicks contractions	Irregular contractions of the uterus that may be felt by the mother in preparation for labor
Fetal presentation	The position of the baby's body parts in relation to the birth canal, such as head-down (cephalic) or breech
Fetal weight	Estimation of the baby's weight through ultrasound measurements
Fetal position	The orientation of the fetus within the uterus, typically described in terms of engagement and station
Placental position	The location of the placenta within the uterus, which can affect delivery options and complications

extracting shape features include stacked convolutional layers, a max-pooling layer, and an MLP layer. The weights and biases in the MLP layer are learned during training using the backpropagation algorithm [37]. The inputs of three blocks can then be represented as $\phi_1 = (H;\Theta_1)$, $\phi_2 = (H;\Theta_1,\Theta_2)$ and $\phi_2 = (H;\Theta_1,\Theta_2,\Theta_3)$. For each input φ_n (where, n = 1,2,3), the corresponding parameter vector ψ_n defines the output of the n^{th} HDR as a high-level features in the density map $(L2(G_n(\phi_n;\psi_n)))$.

Based on density feature maps in Eq. (2), we construct a probability vector, denoted as $P_v \in D^{K\times 1}$, which estimates the likelihood of each organ's shape being representative of network structures. The formulation of P_v is based on this process:

$$P_v = Softmax\left(L2(G_n(\phi_n;\psi_n))\left[1 - N_Z\left[\sum_{k=1}^{K} S_{p,q}\right]\right]\right) \quad (4)$$

where N_z is the normalize and $S_{p,q} = \|s_p - s_q\|_{M_1}$ represents the distance matrix of shape features (p,q).

(iii) **Appearance and Shape-Based DCNN Classifier** Based on the architecture of CNN [38], we apply the DCNN model to combine the features of complex head location and obtain the segmentation map of an accurate fetal head without loss. The appearance and shape features of complex head locations are fed as input to the DCNN for training. Both features are employed to achieve the final segmentation result of the fetal head. Following the convolution layer, nonlinear feature maps are obtained by applying a nonlinear activation function. The DCNN architecture uses the simplest nonlinearity, which is ReLU, and has been demonstrated to be highly effective for segmentation tasks. Transpose convolutional layers are used to increase the spatial resolution of the feature maps. According to the parameters, the input

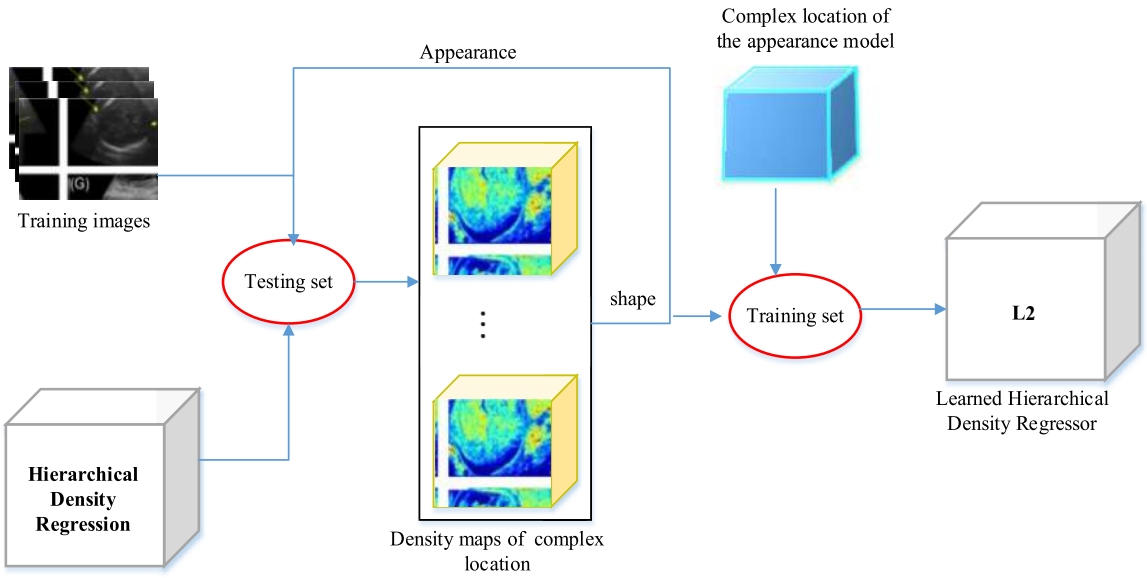


Fig. 4 Framework of hierarchical density regression (L2)

feature map is reduced in size through the pooling layer. A FC layer is employed to transform the feature maps into output class probabilities. This layer uses softmax nonlinearity to convert the 2D feature maps into a 1D feature vector.

The appearance and shape features ($L1(f(H;\Theta))$,$L2\left(G_n\left(\phi_n;\psi_n\right)\right)$) in the feature map ($F_{map}$) are represented as neurons, expressed as:

$$F_{mj} = B_m + \sum_j Y_{mj} * P_j \tag{5}$$

where * denotes the convolutional operation, B_m, P_j and Y_{mj} and represent the bias, input plane, and convolutional kernel, respectively.

ReLUfunctions (0, I_m) result in the computation of the non-linear activation map as $f(m) = MAX\left(0, I_m\right)$. For each activation in a ma A_m, the pooling layer computes the output P(L1, L2) with a size of PL × PL as follows:

$$P(L1, L2) = MAX\left(A_m(L1 + PL, L2 + PL)\right) \tag{6}$$

Although more complex methods may improve performance, we utilize straightforward technique based on DCNN model to produce segmentation maps from point annotations in this research. Set $M^{segmentation} \in [0, 1]^{h \times w}$ for generating the segmentation maps where the height and width of the image are represented as h and w, respectively, expressed as:

$$M^{Segmentaion}(a) = \sum_{k=1}^{K} \gamma\left(a - a_k\right) * G_x(a) \tag{7}$$

where the matrix G_x (a) is a square matrix of size x × x that has an a-centered core. In matrix $M^{Segmentation}$, the pixels corresponding to the foreground and background regions are denoted by ones and zeros, respectively. To ensure that crucial contextual details are not lost, we chose the value of x to be 25 for the experiments. This ensured that the segmentation map had more pixels characterizing a particular fetal head in an image than the density map.

The output target class of fetal head in the appearance and shape features ($L1(f(H;\Theta))$,$L2\left(G_n\left(\phi_n;\psi_n\right)\right)$) are trained together to minimize the combined loss function, expressed as:

$$C_{loss}(L1, L2) = loss(\Theta) + \sum_{n=1}^{3} \beta_n loss_n\left(\Theta_1, ..., \Theta_n, \psi_n\right) + \mu\left(\|\Theta\|^2 + \sum_{n=1}^{3} \|\Theta_n\|^2\right), n = 1,2,3, \tag{8}$$

where the loss function $loss(\Theta)$ calculates the average mean square errors (MSE) between the original US image and the estimated density map obtained through complex density regression (L1). Similarly, the loss function $\beta_n\left(\Theta_1, ..., \Theta_n\psi_n\right)$ calculates the average MSE between the high-level segmentation map generated by the DCNN and the low-level density map calculated by the n-th HDR (L2). Skip connections allow the network to combine low-level and high-level features by skipping over some layers and connecting to later layers in the network. The supervision strength under the n-th DCNN is regulated by the parameter β_n, which ranges from 0 to 1. The loss function $\underline{C_{loss}}$ is computed through the use of momentum stochastic gradient descent (SGD) technique [39]. The parameter n is updated during the i-th iteration, which is denoted by $\Theta_n^{(i)}$. The momentum parameter λ determines the impact of the previous iteration's outcome, while the learning rate η regulates the speed of parameter updating. As $loss_n$ takes values from 1 to n, the gradient with respect to the model parameters Θ_n is solely dependent on Θ_n and Θ_z. The framework of DCNN classifier is shown in Fig. 5.

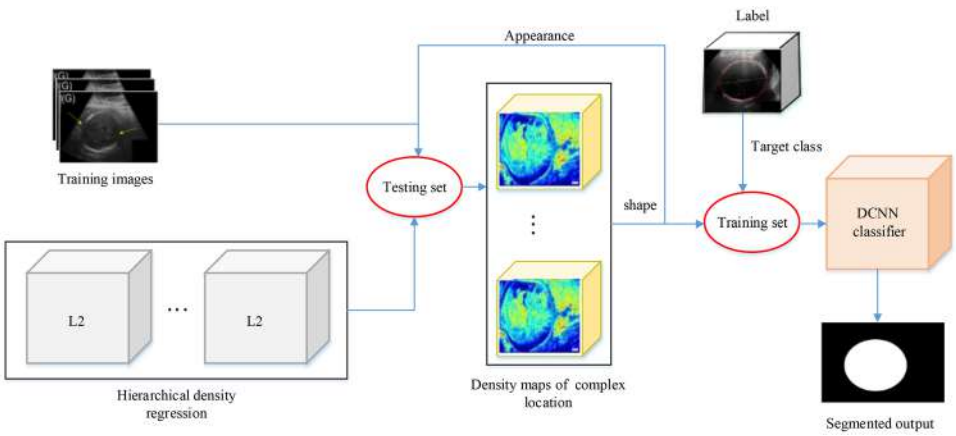


Fig. 5 Framework of DCNN segmentation

Phase 3: Fetal Head Measurements

This phase involves post-processing and HC estimation steps to enhance the accuracy of fitting parameters of FH profiles for HC estimation.

Post-Processing

Since the segmented results are still prone to incomplete problems like blurred FH, different artifact are mistakenly segmented as FH. Therefore, this work employed post-processing operations using contrast enhancement filtering with a morphological operation model to smooth the region and remove unnecessary artifacts from the segmentation results. The operations include area opening, dilation, and filling small gaps. The approach of combining bottom hat (BT) and top hat (TT) contrast enhancement filters (EF) is utilized to effectively distinguish the foreground region from the background in fetal US imaging (FI) [40], expressed as:

$$EF = FI + TT - BT \tag{9}$$

(i) **Area opening operation:** Area opening can be useful in FH post-processing because it can eliminate small regions that may have been falsely included in the segmentation due to noise or other artifacts in the image. Consider r be the number of target classes of FH, A be the average, M_i and M_j represent the major and minor axis lengths of the FH, respectively. The pixel value P_v is expressed as:

$$P_v = R_d\left(mean\left(A\_M_i, A\_M_j\right)\right) \tag{10}$$

where is the round of radii values.

(ii) **Dilation operation:** Dilation is a morphological operation used to expand the boundaries of the segmented FH filling in small gaps and smoothing the edges. Consider M_{min} and M_{max} be the minimum value of major axis length and maximum value of the minor axis length of the FH in the fetal US image, expressed as:

$$Dl = R_d\left(MIN\left(M_{\min}, M_{\max}\right)\right) \tag{11}$$

(iii) **Fill hole operation:** The fill-hole operation is a morphological operation that is used to fill small gaps that may be present in the segmented FH. These gaps may occur due to the presence of noise or other artifacts in the fetal US image. Thus, the post-processing steps obtain a smoothed segmented result, leading to improved quality outcomes for the image and thereby reducing the issues of overlapping and over-segmentation.

Head Circumference Estimation

After post-processing, the HC can be estimated from the FH points in the US images. Numerous approaches to handling the HC have been proposed, but most of them assume that the FH has an irregular shape, even though FH profiles have slight deviations despite their irregular appearance. To enhance the accuracy of fitting parameters for FH profiles, the elliptic model is employed, as it can better represent the shape and circumference measurements of FH [41]. This paper applies the robust ellipse fitting-based least square (REFLS) method based on elliptic hypothesis to avoid the effect caused by incomplete ellipse problems and determine the ellipse parameters such as orientation angle, semi-axis, center coordinates, and inner and outer ellipse scaling parameters for HC estimation. The REFLS method is used to estimate the head circumference of an individual from an US image. The method involves fitting an ellipse to the outline of the head in the image.

For the purpose of the sliced HC-point, we considered them as ellipses and utilized the REFLS algorithm to determine the best fitting points. To prevent the trivial solution, set $\left[E_1, E_2, E_3, E_4, E_5, E_6\right]^T = 0_{[6\times1]}$, the constraint $E_6 = -1$ is imposed. It should be noted that the value of E_6 is not influenced by the edge point (a,b), expressed as:

The principle of least square in the HC data points are expressed as:

$$J = \left(a^T \overline{W} a\right)^{-1}\left(a^T \overline{W} C\right) \tag{12}$$

where C is the constant value and $\overline{W}$ denotes the weight matrix equivalent to the edge points.

According to the model parameters, we introduced the REFLS approach with robust weight function W_f to increase the accuracy of the ellipse parameter solution as follows:

$$W_f = \begin{cases} 1 & \left|\overline{e_f}\right| \le t_o \\ \frac{t_0\left(t_1 - \left|\overline{e_f}\right|\right)}{\left|\overline{e_f}\right|\left(t_1 - t_0\right)} & t_0 \le \left|\overline{e_f}\right| \le t_1 \\ 0 & \left|\overline{e_f}\right| \ge t_1 \end{cases} \tag{13}$$

where the residual error of the measurement is denoted by $\overline{e_f}$, while the standard deviation of the measurement error is represented by σ_0. The modulation coefficients for the robust threshold are t_0 and t_1, and the standardized residuals are denoted by $\left|\overline{e_f}\right| = \frac{|e_f|}{\sigma_0}$.

By utilizing elliptic coefficients such as the two semi-axes (s_1, s_2) (i.e., half long and half-short axis of the ellipse), orientation (ϕ), center coordinates (x_0, y_0), and the inner and outer ellipses (I_E, O_E), the standard equations for the HC measurements fitting in the ellipse can be achieved using the following formulas:

$$I_E = \frac{\left[(a-x_0)\cos\varphi + (b-y_0)\sin\varphi\right]^2}{s_{1,1}^2} + \frac{\left[-\left((a-x_0)\sin\varphi + (b-y_0)\cos\varphi\right)\right]^2}{s_{2,1}^2} = 1 \tag{14}$$

$$O_E = \frac{\left[(a-x_0)\cos\varphi + (b-y_0)\sin\varphi\right]^2}{s_{1,2}^2} + \frac{\left[-\left((a-x_0)\sin\varphi + (b-y_0)\cos\varphi\right)\right]^2}{s_{2,2}^2} = 1 \tag{15}$$

$$E_{coeff} = 4(s_1+s_2) - 4\left[4 - \pi + \frac{0.1218(s_1-s_2)^2}{(s_1+s_2)^2 + 2.8s_1s_2}\right]\frac{s_1s_2}{s_1+s_2} \tag{16}$$

$$HC_E = \frac{E_{coeff}}{\pi}(I_R + O_E) \tag{17}$$

Experiments and Evaluation

In this section, we evaluate the efficacy of our DR-ASPnet approach through a series of experiments. We first provide an overview of the implementation details, hyper-parameters, and evaluation metrics. Next, we present visual comparison results for the proposed and existing methods with other state-of-the-art methods. Finally, we report our findings on the HC18 dataset and provide quantitative results in terms of correlation and regression analysis.

Implementation

This section outlines the implementation process of the DR-ASPnet model in terms of hardware, hyper parameters, and evaluation metrics.

Hardware

The experiments in this study were conducted using a graphics workstation equipped with an Intel(R) Core(TM) i7-3840QM CPU and 64 GB RAM. The widely used public PYTHON library based on Tensorflow 2.6.0 and Keras 2.4.0 were selected for the evaluation of the DR-ASPnet framework.

Hyper Parameters

Some of the hyper-parameter configurations have been described in the "Data Description" section. The network configuration parameters are shown in Table 4.

Table 4 Configuration parameters

Parameters	Values
Epochs	20 for each iteration
Epoch time	65 min
Total running time	2 h
Optimizer	SGD [39]
Learning rate	0.001
Batch size	32
Drop our ratio	0.5

Evaluation metrics

We considered seven metrics, such as DSC, Handoff distance (HD), mean ntersection over union (mIoU), coefficient of variation (CV), difference (DF), AD, and mean absolute error (MAE) to evaluate the efficiency of the DR-ASPnet model.

(i) **DSC** [42]: It is also known as the dice coefficient or dice index, which is a statistical measure used to evaluate the similarity between two areas (segmentation area and measurement area). The DSC ranges between 0 and 1, with a value of 1 indicating a perfect overlap or complete agreement between the two areas, and a value of 0 indicating no overlap or disagreement.

(ii) **HD** [43]: It is the maximum distance calculated by Euclidean distance, where the segmentation and measurements are defined by area contours.

(iii) **MIoU** [44]: The mean IoU computes the average IoU across multiple segmented areas (FH and background measurement). It provides an overall measure between the automated segmentation algorithms aligns with the ground truth annotations. A higher mean IoU value indicates better segmentation performance, suggesting a higher degree of overlap and accuracy between the automated and manual segmentations.

(iv) **CV** [45]: This is a statistical measure used to assess the relative variability of a set of measurements (standard deviation and mean value differences between segmentation area and measurement area). A lower CV value indicates less variability and greater reliability in the segmentation or measurement outcomes, suggesting more consistent and accurate results. Conversely, a higher CV value implies greater variability and less reliability, indicating potential inconsistencies or errors in the segmentation or measurement process.

(v) **DF** [28]: The DF and AD metrics are often employed to assess the degree of variation or deviation between

the manual annotations and the automated segmentation results for HC measurement. Smaller absolute differences indicate closer agreement and higher accuracy, while larger absolute differences suggest greater discrepancies or errors.

(vi) **MAE** [46]: This metric can be utilized to evaluate the dissimilarity between the predicted and measured HC for the number of samples. It is calculated as the average of the absolute differences between the corresponding values. A smaller MAE value indicates a closer agreement between the predicted and measured results, suggesting higher accuracy. Conversely, a larger MAE value indicates greater discrepancies or errors between the two sets of values, indicating lower accuracy.

Experimental Results and Analysis

This section presents the performance comparison results of the DR-ASPnet approach based on four sub-sections: visual comparison results, comparison between different trimesters, comparison with existing state-of-the-art methods, and results on correlation and regression analysis for qualitative and quantitative evaluation.

Visual Comparison Results

The performance results of the DR-ASPnet approach are compared with existing methods such as mask regression CNN (Mask R-CNN) [47], Mini Link-Net [22], scale attention feature pyramid etwork (SAFNet) [48], scale attention pyramid deep neural network (SAPNet) [49], R-CNN [25], DAG V-Net [30], RR-CNN [29], distance-field regression version of Mask-RCNN (Mask-R2 CNN) [50], ensemble transfer learning model (ETLM) [51], and Deep-Link-Net [31] for segmentation accuracy and HC estimation accuracy fetal in US images.

To statistically analyze the significant differences in FH measurements between positive and negative points, the Wilcoxon signed-rank test [52] is a non-parametric test that compares the medians of the paired measurements and determines if there is a statistically significant difference. The paired measurements of FH sizes are gathered to calculate positive and negative points for the proposed DR-ASPnet method and existing methods such as Mask R-CNN, Mini Link-Net, SAFNet, SAPNet, R-CNN, DAGV-Net, RR-CNN, Mask-R2 CNN, ETLM, and Deep-Link-Net. Ensure that each pair corresponds to the same fetus. Then state the null and alternative hypotheses between positive and negative points. We performed a Wilcoxon signed-rank test to examine the differences in FH measurements between positive and negative points. The sample consists of N participants. The test statistic was calculated as X, resulting in a p-value of Y. As the p-value was less than the significance level of 0.05, we reject the null hypothesis and conclude that there is a significant difference in FH measurements between positive and negative points. From the analysis, the best results are determined by a Wilcoxon signed-rank test that indicates medium, large, or small effect sizes for the proposed and existing methods, as shown in Figs. 6, and 7.

The qualitative comparisons for the automatic FH segmentation and HC estimation between the DR-ASPnet model and existing methods are shown in Figs. 6, and 7. The qualitative findings reveal that the DR-ASPnet network outperforms other methods in handling incomplete regions of the head while generating smooth segmentations for three different trimesters, as demonstrated in Fig. 6. Furthermore, saliency maps highlight inaccuracies in estimations by regression models, which are misled by high-intensity pixels above the head, resulting in lower predicted HC values. Figure 7 illustrates that the DR-ASPnet model, in certain cases, relies on a significant number of contour pixels to estimate the HC and effectively extracts appearance and shape features from the head contour of US images.

Comparison Between Different Trimesters of US Images

The US images captured during the first trimester are often riddled with noise, hazy regions, and unclear skull borders. Existing regression-based CNN models can lose their way in some areas, confusing other outlines with skulls. During the second trimester, there is a sudden jump in the fetal skull boundary, causing segmentation confusion. In the third trimester, black templates covering other fetal data in the dataset make the discontinuous skull patches in the US image more pronounced and irregular. We encountered these difficulties while testing the DR-ASPnet method.

The evaluation metrics analyze FH measurements across trimesters of US images to investigate if there are significant differences in various metrics (DSC, DF, AD, HD, MAE, and mIoU). This analysis employed an analysis of variance (ANOVA) test [53] to compare the means of the measurements between the trimesters for the proposed and existing methods. A dataset consisting of FH measurements (e.g., head circumference) was collected from US images. The data were categorized according to the trimesters: first trimester (T1), second trimester (T2), and third trimester (T3). The ANOVA test was selected to determine if there were significant differences in the means of the metrics among the trimesters. The ANOVA test revealed a significant difference in mean HC among the trimesters (F-value = [F-value], p-value < 0.05). The statistical analysis using ANOVA revealed significant differences in all the measured FH metrics across the trimesters. These findings suggest that the development of FH dimensions varies significantly as pregnancy progresses.

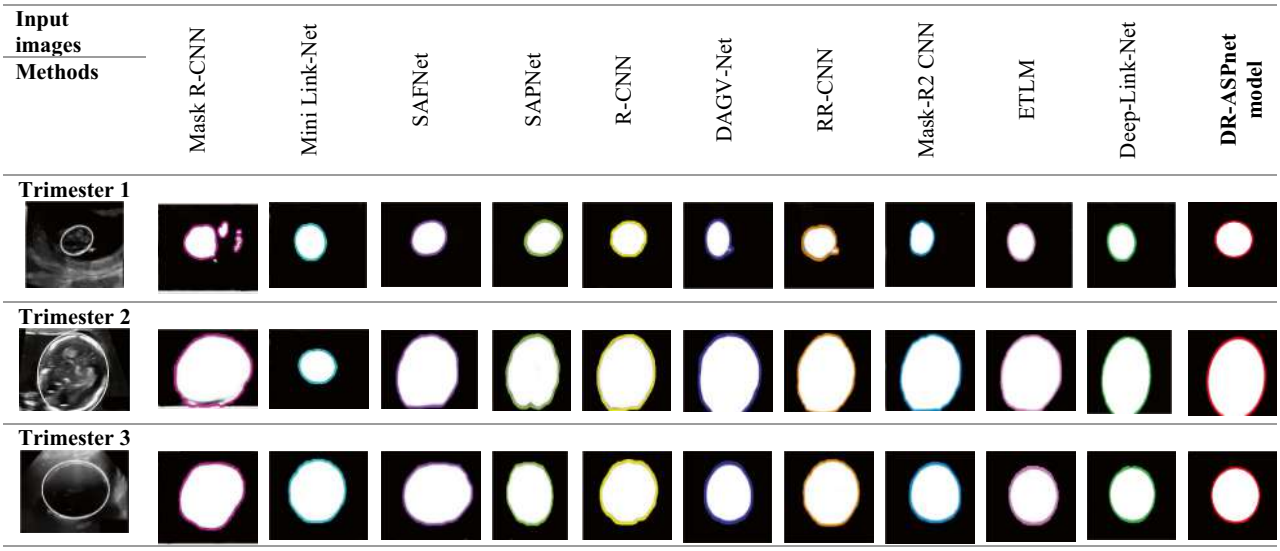


Fig. 6 Visual comparison results for the automatic FH segmentation: first row represents the sample input images from three types of trimesters on HC18 dataset; first column represents the segmentation outputs for the DR-ASPnet method and exiting methods, where the segmented fetal head region after discarding numerous volumes and different colors indicates contours for each comparison models. The best results are shown by the DR-ASPnet model.

Table 5 presents the performance of the DR-ASPnet model after post-processing and elliptical fitting on HC18 testing sets. The DR-ASPnet approach showed excellent accuracy, with a DSC of over 98.86% in the third trimester. In this study, the graphics workstation used achieved an average running time of 5.4 ms for the DR-ASPnet automated fetal HC segmentation and measurement when tested on US images.

Figure 8 indicates that the DR-ASPnet FH segmentation model is accurately fitting the data with less training and validation loss. The learning curves plot the DR-ASPnet model loss on the *y*-axis and the number of training epochs on the *x*-axis. The training loss refers to the error between the model's predicted output and the measured labels during the training phase. The validation loss, on the other hand, measures the error on a separate validation set for the model's generalization performance. Using the DR-ASPnet model, both the training and validation losses are decreasing and stabilizing at low values, which may indicate that the model is learning the underlying patterns effectively.

Comparison with Existing State-of-the-Art Methods

The qualitative evaluation of the DR-ASPnet model is compared with existing state-of-the-art methods such as Mask R-CNN, Mini Link-Net, SAFNet, SAPNet, R-CNN, DAGV-Net, RR-CNN, Mask-R2 CNN, ETLM, and Deep-Link-Net in terms of DSC, DF, AD, HD, MAE, and mIoU metrics.

The evaluation metrics analyze the segmentation accuracy and HC estimation accuracy in US image data and investigate if there are significant differences in these metrics. The analysis will employ statistical tests to compare the performance of segmentation and HC estimation across different methods of the proposed DR-ASPnet model and the existing Mask R-CNN, Mini Link-Net, SAFNet, SAPNet, R-CNN, DAGV-Net, RR-CNN, Mask-R2 CNN, ETLM, and Deep-Link-Net. A dataset of US images, along with corresponding ground truth annotations and HC measurements, was collected. Various algorithms were applied to obtain the segmented area. HC estimation algorithms were then used to estimate the HC from the segmented area. The segmentation accuracy and HC estimation accuracy were evaluated using appropriate metrics (DSC, DF, AD, HD, MAE, and mIoU). Statistical tests are conducted to determine if there are significant differences in the performance of segmentation and HC estimation methods. The statistical test revealed a significant difference in the HC estimation accuracy and segmentation accuracy between the proposed and existing methods (p-value < 0.05). Pairwise comparison tests are performed to identify the specific differences among the methods. The statistical analysis indicates significant differences in both segmentation accuracy and HC estimation accuracy across the tested methods. These findings suggest variations in the performance of different algorithms for segmenting area and estimating HC in US image data, as shown in Table 6.

From Table 7, the evaluation result of the DR-ASPnet model achieves better results than other methods in terms of FH segmentation accuracy and HC estimation accuracy

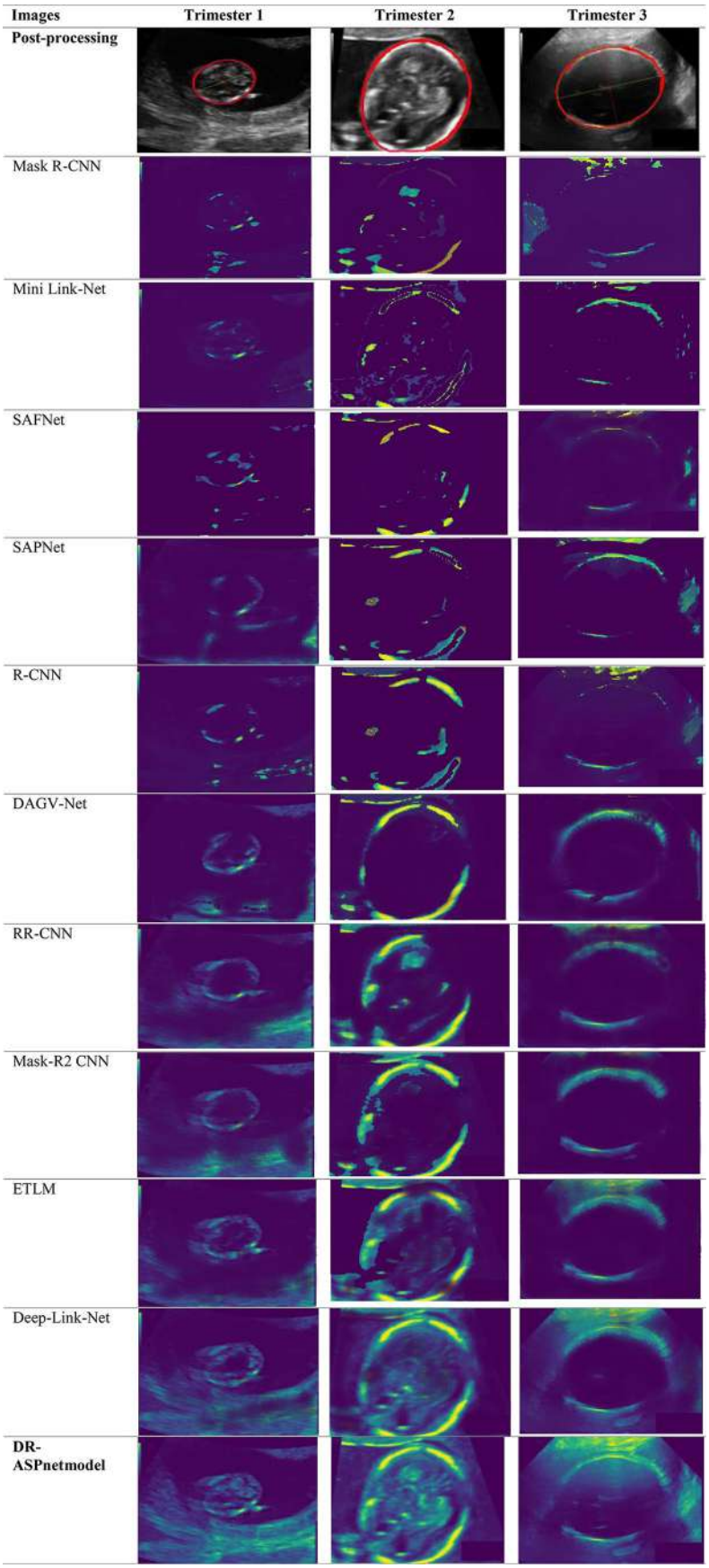


Fig. 7 Saliency map results for the HC estimation values: first row represents the final FH contours for three trimesters, where the red line indicates the boundary of the segmented image; first column represents the HC estimation outputs for the DR-ASPnet method and exiting methods. The best results are shown in bold font. The yellow color denotes positive points, and blue color denotes negative points

Table 5 Performance results of DR-ASPnet model segmentation and measurements on HC18 dataset in terms of evaluation metrics

Trimesters of US images	Evaluation metrics					
	DSC (%)	DF(mm)	AD (mm)	HD(mm)	MAE (mm)	mIoU(%)
First trimester	98.1	0.32	1.527	0.875	1.90	98.03
Second trimester	98.60	–0.122	1.875	1.254	1.92	98.56
Third trimester	98.86	–0.347	2.536	2.364	1.94	98.71

metrics. The regression models such as Mask R-CNN, R-CNN, RR-CNN, and Mask R2-CNN are sensitive to variations in image quality, and their performance may degrade when applied to low-quality images with artifacts, noise encountered in clinical practices. Such models are trained on one set of measurement techniques due to variability in fetal HC measurement, leading to reduced accuracy of the predictions. The list of other remaining methods, such as the ETLM model, may not fully account for image variability, resulting in potentially inaccurate predictions. Then, the Mini Link-Net model results in reduced capacity for capturing complex and fine-grained features in US images. These results lower segmentation accuracy and HC estimation accuracy. Among these facts, the DR-ASPnet model obtains a higher accuracy of 98.86% and mIoU of 98.73% than other methods.

The advantages of the proposed DR-ASPnet model are as follows: (a) it is specifically designed for segmentation and estimation tasks, making it well-suited for accurately delineating fetal US images. The model's deep architecture learns complex features from the data, leading to improved segmentation accuracy, which is reflected in higher DSC scores. (b) The combination of deep CNNs and robust ellipse fitting leads to enhanced segmentation accuracy. Deep CNNs are adept at learning intricate features from medical images, while robust ellipse fitting helps to accurately delineate the shape of fetal structures, leading to improved segmentation. (c) The integration of shape priors and robust ellipse fitting could potentially reduce false positives and negatives in segmentation. The method helps the algorithm make more informed decisions in the segmentation mask, which contributes to better estimation results.

Ablation Experiment

In this part, we validate the effectiveness of each key component used in our proposed DR-ASPnet model. The ablation study involves an ablation experiment on the structure of the baseline regression networks (Mask R-CNN, R-CNN, RR-CNN, and Mask-R2 CNN). All ablation experiments were conducted on the HC18 dataset.

To prove the effectiveness of the proposed DR-ASPnet model, we report the quantitative comparison results of our model with other related architectures. The results are presented in Table 8. R-CNN combines region proposal generation with CNN-based feature extraction. While R-CNN can be adapted for segmentation tasks, it may not be the most efficient choice for real-time or densely segmented medical imaging like fetal US due to its time-consuming region proposal step. Mask R-CNN provides pixel-level segmentation masks for each detected object. This can be valuable in applications like fetal US, where precise delineation of fetal structures is important. Recurrent mask (Mask R-CNN and Mask-R2 CNN) connections allow the network to capture temporal dependencies, which might be useful in fetal US images where structures change over time. Residual connections help alleviate the vanishing gradient problem and enable the training of deeper networks. From the overall

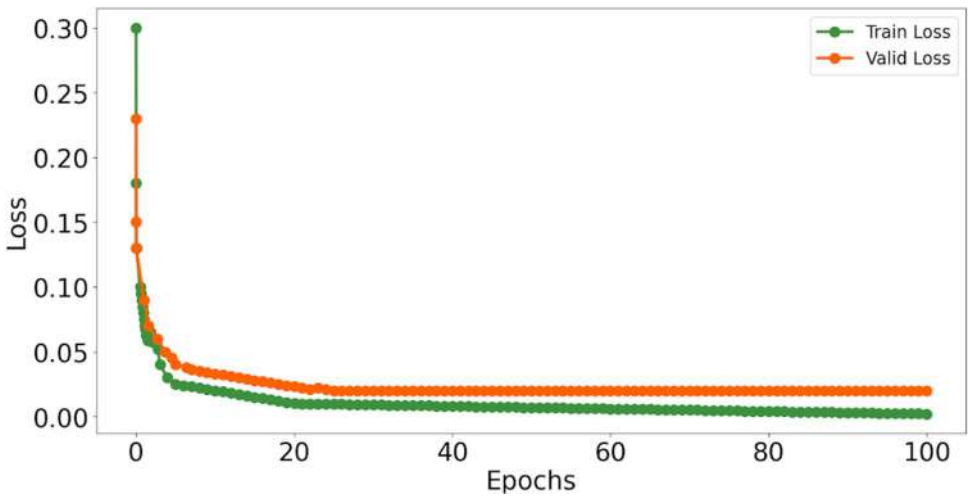


Fig. 8 Training loss and validation loss of the DR-ASPnet model

Table 6 Statistical significant test results in terms of various metrics

Models	Metrics											
	Mask R-CNN	Mini Link-Net	SAFNet	SAPNet	R-CNN	DAGV-Net	RR-CNN	Mask-R2 CNN	ETLM	Deep-Link-Net	DR-ASPnet	*p*-value
DSC	0.977	0.968	0.981	0.955	0.945	0.979	0.977	0.986	0.972	0.936	0.988	0.001
DF	1.48	1.12	1.31	1.13	0.16	0.21	0.11	0.31	1.00	2.11	0.133	0.021
AD	2.44	2.14	2.11	2.01	4.53	1.774	1.92	1.94	1.952	2.11	1.677	0.008
HD	1.44	1.77	1.33	1.32	1.34	1.36	1.44	1.56	1.76	1.892	1.223	0.003
MAE	2.42	2.32	2.11	2.10	1.98	1.995	1.97	1.96	1.95	1.9	1.875	0.012
MIoU	93.99	93.8	95.44	95.06	96.55	96.06	97.44	97.56	98.53	98.6	98.75	0.004

Table 7 Performance comparison results of segmentation accuracy and HC estimation accuracy

Methods	Evaluation metrics					
	DSC (%) ± std	**DF(mm) ± std**	**AD (mm) ± std**	**HD(mm) ± std**	**MAE (mm) ± std**	**mIoU(%) ± std**
Mask R-CNN [47]	97.7 ± 1.3	1.49 ± 2.8	2.4 ± 2.2	1.4 ± 0.82	-	-
Mini Link-Net [22]	96.8 ± 2.9	1.13 ± 2.7	2.1 ± 1.9	1.7 ± 1.4	-	-
SAFNet [48]	98.1 ± 4.0	1.3 ± 2.9	-	1.3 ± 0.8	-	-
SAPNet [49]	95.57 ± 2.3	-	-	-	-	95.06 ± 2.2
R-CNN [25]	-	-	4.5 ± 4.3	-	-	-
DAGV-Net [30]	97.9 ± 1.3	0.16 ± 2.5	1.77 ± 1.7	1.3 ± 0.8	-	-
RR-CNN [29]	97.7 ± 1.3	0.21 ± 2.6	1.9 ± 1.7	1.3 ± 0.7	1.9 ± 1.7	-
Mask-R2 CNN [50]	98.6 ± 0.9	-	-		1.95 ± 1.3	-
ETLM [51]	97.2 ± 1.4	-	-	1.77 ± 0.9	-	98.53 ± 2.3
Deep-Link-Net [31]	93.6 ± 2.8	1.03 ± 3.8	2.1 ± 2.9	1.89 ± 6.3	-	-
DR-ASPnet model	**98.86 ± 0.8**	**0.13 ± 2.2**	**1.67 ± 1.5**	**1.22 ± 0.77**	**1.87 ± 1.23**	**98.73 ± 2.1**

The (mm) results are mean and ± standard deviation (std);—indicates not available and the bold indicates best results

Table 8 Ablation study on different architectures (where the bold indicates best results)

	Mean DSC	Mean HD	Mean PA
R-CNN	0.821 ± 0.115	3.514 ± 0.862	0.916 ± 0.047
RR-CNN	0.866 ± 0.115	2.452 ± 0.895	0.945 ± 0.042
Mask R-CNN	0.883 ± 0.158	2.321 ± 0.887	0.962 ± 0.041
Mask-R2 CNN	0.945 ± 0.109	1.390 ± 0.874	0.926 ± 0.040
DR-ASPnet	**0.986 ± 0.081**	**1.230 ± 0.864**	**0.987 ± 0.039**

results, the proposed DR-ASPnet model achieves the best performance among other models in terms of DSC, HD, and pixel-level accuracy (PA) metrics.

Results on Correlation and Regression Analysis

In this section, we evaluate the fetal biometric parameter of HC using the DR-ASPnet method on the training and testing sets of different trimesters under the HC18 dataset. A Bland–Altman plot, a scatter plot, and a box plot were employed to perform correlation and regression analysis between the measured and predicted results.

The Bland–Altman plots for the fetal biometric parameter HC, which is generated for clinical validation using DR-ASPnet method on the training and testing sets of different trimesters, are displayed in Fig. 9. In the context of fetal HC parameter, each data point on the Bland–Altman plot represents the difference between the measured and predicted HC values on the training and testing sets of the HC18 dataset. In addition, Bland–Altman plots often include reference lines to indicate the mean difference and the limits of agreement. The mean difference line is a horizontal dashed line represents the average difference between the measured and predicted HC values. Scatter plot compares the measured and predicted fetal HC values for the DR-ASPnet model on the training and testing sets of different trimesters, which is shown in Fig. 10. Each data point on the box plot of absolute error values represents the absolute difference between the measured fetal HC and the true value and is used as the data by the DR-ASPnetmodel, which is shown in Fig. 10.

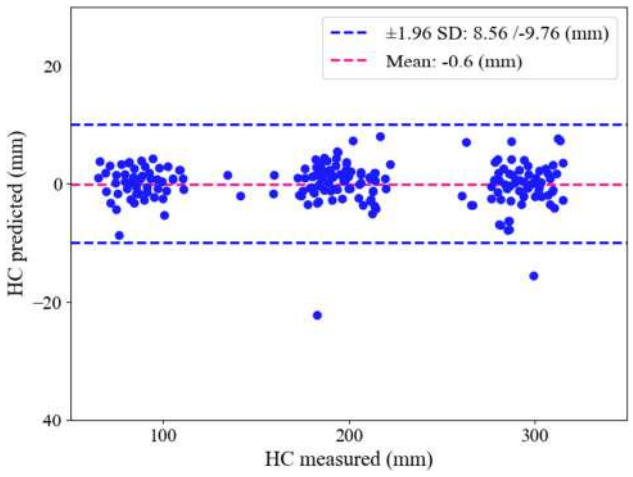


Fig. 9 Bland–Altman plots for fetal HC parameter. The differences are often plotted as positive and negative values (in mm)

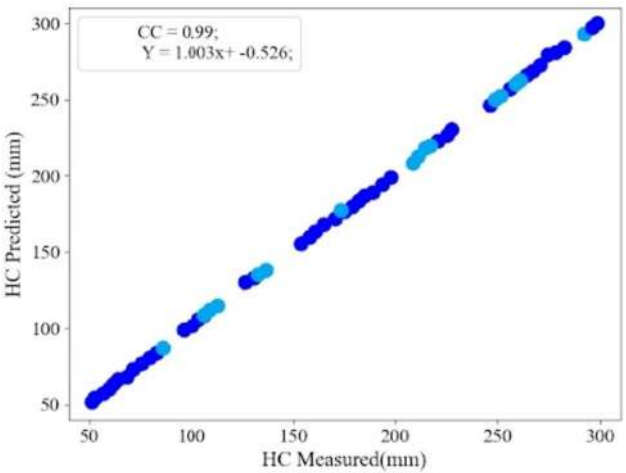


Fig. 10 Scatter plots for fetal HC parameter

From Figs. 9, 10, and 11, the overall plotted results show that the DR-ASPnet model obtains a strong correlation between measured and predicted HC values. It provides visual insights into the agreement between the measured and predicted HC values for testing and training sets. Table 9 shows the correlation and regression fit values for the HC18 dataset. The statistical analysis of the relationship across the three trimesters revealed variations based on the modeling of HC absolute error. When HC absolute error was represented as a z-score, a significant association with trimester classes emerged in the first, second, and third trimesters ($p = 0.005$ and $p < 0.0001$, respectively). However, when HC absolute error was assessed using percentiles (3 categories), trimester classes displayed no significant association with HC absolute error in the second trimester ($p = 0.42$), while the association remained significant in the third trimester ($p < 0.0001$). The distributions of HC absolute error for the second and third trimesters within each trimester category are illustrated in Fig. 11.

Table 9 Correlation and regression values on the training and testing sets of different trimesters. (CC denotes the correlation analysis, Y represents the regression analysis, AE is the absolute error, where $Y = \mathrm{p}\,\mathrm{x} + \mathrm{q}$, CC = 1, $p = 1$ and $q = 0$)

Trimesters of US images	HC
Training set (999 images)	CC = 1; $p = 1.004$; $q = -0.566$
Testing set (335 images)	CC = 0.99; $p = 1.003$; $q = -0.526$

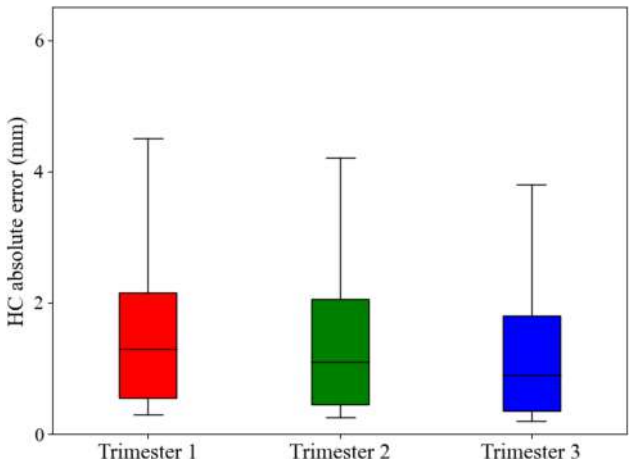
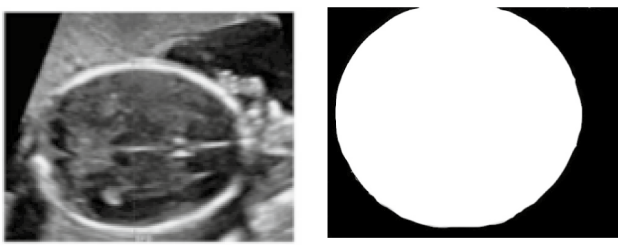


Fig. 11 Box plot results for fetal HC parameter

Discussion

This section presents a deep analysis of existing state-of-the-art methods and the proposed method, which lies in the broader context of the relevant literature comparison results using US images. At first, the CAD system observes the initial knowledge of ultrasound scanning and GA to locate the FH. Further, the center line of the fetal skull is detected using the phase symmetry approach [27]. The first GA was then calculated using the crown-rump length (CRL) measurement. Here, the center and margin of the fetal skull are found using three HC measurement methods. Last but not least, the HC from the detector output was measured using an elliptical fit [28]. Further, HC measurement work has been extended using deep learning (DL) techniques. Then, some mislabeled US images are smoothly polished using ellipse fitting during post-processing [21]. Two edge-localization tasks were used [29]: (1) location and center part of the FH are determined by region CNN structure trained based on transfer learning; (2) HC is accurately delineated using R-CNN trained based on distance-field. However, HC pixel position poses challenges to convergence performance during training. Hence, the obtained results achieve the highest DSC and the lowest mean absolute difference (MAD) error [21, 27–29].



(a) (b)

Fig. 12 Failure results of the proposed model: **a** US image, **b** worst performance result of segmented and fitted ellipse

Following the common processes of edge detection, morphological processing, and ellipse fitting, the DAG V-net model used two trained models: deep supervision and attention gate mechanisms for automatic FH segmentation. However, it is important to note that the DAG V-Net model has some limitations, including a narrow scope of cases used for training and lower accuracy during the first trimester of pregnancy [30]. From the evaluation results, the Deep-Link-Net-based HC measurement achieves better efficiency in terms of MAD and Handoff metrics [31]. Four source datasets (unseen) and target datasets (seen) were adopted using the synergistic learning method [32] to solve the intra-class inconsistency segmentation [30–32]. Among these facts, the evaluation results from these methods [21, 27–32] have suffered from many problems, such as the fact that whole skull boundaries are difficult to extract from US images, and are high time-consuming due to the absence of specialized knowledge about different image conditions, the segmentation process cannot fulfill high accuracy for three trimesters, and the training parameters are limited to conduct the experiments when simulated on real-time data. To solve the existing state-of-the-art methods, the proposed DR-ASPnet model qualitative findings outperform existing methods in handling incomplete regions of the head while generating smooth segmentations for three different trimesters. Furthermore, in certain cases, it relies on a significant number of contour pixels to estimate the HC and effectively extracts appearance and shape features from the head contour of US images. Further, the DR-ASPnet approach shows excellent accuracy; a DSC of over 98.86% with low training and validation loss values indicates that the model is learning the underlying patterns effectively. Besides, it takes a shorter average running time of 5.4 ms for fetal HC segmentation and measurement when tested on US images.

In the context of the proposed DR-ASPnet model, the worst-case performance indicates that the method performs under the most challenging and difficult conditions. From the worst case performance shown in Fig. 12, the DR-ASPnet model require large and diverse training datasets to generalize well to new, unseen data. However, the training data does not cover a broad spectrum of FH appearances, or if it's not representative of the conditions, the segmentation model might fail to adapt to new cases. The robustness of the ellipse fitting step can also impact segmentation results. However, the DR-ASPnet method relies heavily on the robustness of ellipse fitting, and the initial conditions for fitting the ellipse are not accurate, which can lead to failures in cases where the FH shape deviates significantly from an ellipse. Even though the proposed DR-ASPnet model

obtains potential results, there exist some limitations. (1) In certain cases, fetal structures may be partially obscured or difficult to visualize in US images. This can occur due to maternal obesity and suboptimal imaging conditions. These challenges can affect the accuracy of segmentation and subsequent measurement estimation. (2) Fetal US images can exhibit significant anatomical variability due to gestational age and maternal factors. This variability can pose challenges in accurately segmenting fetal structures and fitting ellipses to estimate measurements. (3) Depending on the healthcare setting, acquiring patient 3D volumes is more time-consuming and resource-intensive compared to obtaining 2D US images. This can be due to cost constraints and the limited availability of specialized equipment.

Conclusion and Future Scope

This paper presents automatic FH segmentation and HC estimation using the DR-ASPnet model with post-processing and robust ellipse fitting in US images. The image pre-processing and augmentation steps were applied to solve the blurred image and variations in different pixel sizes over different trimesters for the training and testing processes. Specifically, by developing three network models, i.e., appearance-based complex density regression, appearance and shape-based hierarchical density regression, and an appearance and shape-based DCNN classifier, our DR-ASPnet model essentially solves the problems of FH segmentation from US images. Next, we employed three morphological operations, i.e., area opening, dilation, and fill hole, to smooth the segmented output in the post-processing. Then, we applied the REFLS method to determine the best fitting parameters from the FH points for HC estimation in US images. Implementation results of the DR-ASPnet model attain higher segmentation accuracy and measurement accuracy with less running time on the publicly available HC18 dataset and also obtain better performance than other methods. In future work, we could plan to investigate the proposed model on other medical datasets to estimate other fetal biometry parameters. Besides, we will focus on developing 3D segmentation algorithms that can handle the inherent challenges of working with volumetric data on the fetal head.

Author Contributions All the authors have participated in writing the manuscript and have revised the final version. All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by GD, SS, AKJ, MS, PS, and MM. The first draft of the manuscript was written by GD, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. Conceptualization: GD, SS. Methodology: GD, AKJ. Formal analysis and investigation: GD, MS, SS. Writing — original draft preparation: GD, PS, MM. Writing — review and editing: GD, AKJ, MS. Supervision: MM.

Data Availability The dataset is public and can be downloaded from https://hc18.grand-challenge.org/.

Declarations

Ethical Approval This article does not contain any studies with human participants and/or animals performed by any of the authors.

Informed Consent There is no informed consent for this study.

Conflict of Interest The authors declare no competing interests.

References

1. Figueras F, Caradeux J, Crispi F, Eixarch E, Peguero A, GratacosE: Diagnosis and surveillance of late-onset fetal growth restriction. American journal of obstetrics and gynecology. 218(2):S790-802, 2018
2. Necas M: The clinical ultrasound report: Guideline for sonographers. Australasian Journal of Ultrasound in Medicine. 21(1):9-23, 2018
3. Matthew J, Skelton E, Day TG, Zimmer VA, Gomez A, Wheeler G, Toussaint N, Liu T, Budd S, Lloyd K, Wright R: Exploring a new paradigm for the fetal anomaly ultrasound scan: Artificial intelligence in real time. Prenatal diagnosis. 42(1):49-59, 2022
4. World Health Organization. WHO antenatal care recommendations for a positive pregnancy experience: maternal and fetal assessment update: imaging ultrasound before 24 weeks of pregnancy.
5. Mamsen LS, Björvang RD, Mucs D, Vinnars MT, Papadogiannakis N, Lindh CH, Andersen CY, Damdimopoulou P: Concentrations of perfluoroalkyl substances (PFASs) in human embryonic and fetal organs from first, second, and third trimester pregnancies. Environment international. 124:482-92, 2019
6. O'Gorman N, Salomon LJ: Fetal biometry to assess the size and growth of the fetus. Best practice & research Clinical obstetrics & gynaecology. 49:3-15, 2018
7. Mandal SK, Ghosh SK, Roy S, Prakash B: Evaluation of fetaltranscerebellar diameter as a sonological parameter for the estimation of fetal gestational age in comparison to biparietal diameter and femur length. age. 6:7, 2019
8. Hammami A, Mazer Zumaeta A, Syngelaki A, Akolekar R, Nicolaides KH: Ultrasonographic estimation of fetal weight: development of new model and assessment of performance of previous models. Ultrasound in obstetrics &gynecology. 52(1):35-43, 2018
9. O'Brien CM, Louise J, Deussen A, Dodd JM: In overweight and obese women, fetal ultrasound biometry accurately predicts newborn measures. Australian and New Zealand Journal of Obstetrics and Gynaecology. 60(1):101-7, 2020
10. Salomon LJ, Alfirevic Z, Da Silva Costa F, Deter RL, Figueras F, Ghi TA, Glanc P, Khalil A, Lee W, Napolitano R, Papageorghiou A: ISUOG Practice Guidelines: ultrasound assessment of fetal biometry and growth. Ultrasound in obstetrics &gynecology. 53(6):715-23, 2019
11. Burgos-Artizzu XP, Coronado-Gutiérrez D, Valenzuela-Alcaraz B, Vellvé K, Eixarch E, Crispi F, Bonet-Carne E, Bennasar M, Gratacos E: Analysis of maturation features in fetal brain ultrasound via artificial intelligence for the estimation of gestational age. American Journal of Obstetrics &Gynecology MFM. 3(6):100462, 2021
12. Sun Y, Yang H, Zhou J, Wang Y: ISSMF: Integrated semantic and spatial information of multi-level features for automatic segmentation in prenatal ultrasound images. Artificial Intelligence in Medicine. 125:102254, 2022

13. AdithyaPC, Sankar R, Moreno WA, Hart S: Trends in fetal monitoring through phonocardiography: Challenges and future directions. Biomedical Signal Processing and Control. 33:289–305, 2017
14. Kiserud T, Piaggio G, Carroli G, Widmer M, Carvalho J, Neerup Jensen L, Giordano D, Cecatti JG, Abdel Aleem H,Talegawkar SA, Benachi A: The World Health Organization fetal growth charts: a multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. PLoS medicine. 14(1):e1002220, 2017
15. Sahli H, Ben Slama A, MouelhiA, Soayeh N, Rachdi R, Sayadi M: A computer-aided method based on geometrical texture features for a precocious detection of fetal Hydrocephalus in ultrasound images. Technology and Health Care. 28(6):643-64, 2020
16. Rajinikanth V, Dey N, Kumar R,Panneerselvam J, Raja NS: Fetal head periphery extraction from ultrasound image using Jaya algorithm and Chan-Vese segmentation. Procedia Computer Science. 152:66-73, 2019
17. Zhang L, Dudley NJ, Lambrou T, Allinson N, Ye X: Automatic image quality assessment and measurement of fetal head in two-dimensional ultrasound image. Journal of Medical Imaging. 4(2):024001, 2017
18. Cerrolaza JJ, Oktay O, Gomez A, Matthew J, Knight C, Kainz B, Rueckert D: Fetal skull segmentation in 3D ultrasound via structured geodesic random forest. In Fetal, Infant and Ophthalmic Medical Image Analysis: International Workshop, FIFI 2017, and 4th International Workshop, OMIA 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 4 (pp. 25–32). Springer International Publishing, 2017
19. Komatsu M, Sakai A, Dozen A, Shozu K, Yasutomi S, Machino H, Asada K, Kaneko S, Hamamoto R: Towards clinical application of artificial intelligence in ultrasound imaging. Biomedicines. 9(7):720, 2021
20. Cerrolaza JJ, Sinclair M, Li Y, Gomez A, Ferrante E, Matthew J, Gupta C, Knight CL, Rueckert D: Deep learning with ultrasound physics for fetal skull segmentation. In2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (pp. 564–567). IEEE, 2018
21. Skeika EL, Da Luz MR, Fernandes BJ, Siqueira HV, De Andrade ML: Convolutional neural network to detect and measure fetal skull circumference in ultrasound imaging. IEEE Access. 8:191519-29, 2020
22. Sobhaninia Z, Emami A, Karimi N, Samavi S: Localization of fetal head in ultrasound images by multiscale view and deep neural networks. In2020 25th International Computer Conference, Computer Society of Iran (CSICC) (pp. 1–5). IEEE, 2020
23. Wu L, Xin Y, Li S, Wang T, Heng PA, Ni D: Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation. In2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017) (pp. 663–666). IEEE, 2017
24. Qiao D, Zulkernine F: Dilated squeeze-and-excitation U-Net for fetal ultrasound image segmentation. In2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) (pp. 1–7). IEEE, 2020
25. Zhang J,Petitjean C, Lopez P, Ainouz S: Direct estimation of fetal head circumference from ultrasound images based on regression CNN. InMedical Imaging with Deep Learning (pp. 914–922). PMLR, 2020
26. Zhao L, Li N, Tan G, Chen J, Li S, Duan M: The End-to-end Fetal Head Circumference Detection and Estimation in Ultrasound Images. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2022
27. Li J, Wang Y, Lei B, Cheng JZ, Qin J, Wang T, Li S, Ni D: Automatic FH circumference measurement in ultrasound using random forest and fast ellipse fitting. IEEE journal of biomedical and health informatics. 22(1):215-23, 2017
28. van den Heuvel TL, de Bruijn D, de Korte CL, Ginneken BV: Automated measurement of fetal head circumference using 2D ultrasound images. PloS one. 13(8):e0200412, 2018
29. Fiorentino MC, Moccia S, Capparuccini M, Giamberini S, Frontoni E: A regression framework to head-circumference delineation from US fetal images. Computer methods and programs in biomedicine. 198:105771, 2021
30. 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. Journal of Digital Imaging. 34:134-48, 2021
31. Amini SM: Head circumference measurement with deep learning approach based on multi-scale ultrasound images. Multimedia Tools and Applications. 81(23):32981-93, 2022
32. Zhou M, Wang C, Lu Y, Qiu R, Zeng R, Zhi D, Jiang X, Ou Z, Wang H, Chen G, Bai J: The segmentation effect of style transfer on FH ultrasound image: a study of multi-source data. Medical & Biological Engineering & Computing. 1–5, 2023
33. Pazinato DV, Stein BV, de Almeida WR, Werneck RD, Júnior PR, Penatti OA, Torres RD, Menezes FH, Rocha A: Pixel-level tissue classification for ultrasound images. IEEE journal of biomedical and health informatics. 20(1):256-67, 2014
34. Medak D, Posilović L, Subašić M, Budimir M, Lončarić S: Automated defect detection from ultrasonic images using deep learning. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. 68(10):3126-34, 2021
35. Korneev S, Narodytska N, Pulina L, Tacchella A, Bjorner N, Sagiv M: Constrained image generation using binarized neural networks with decision procedures. InTheory and Applications of Satisfiability Testing–SAT 2018: 21st International Conference, SAT 2018, Held as Part of the Federated Logic Conference, FloC 2018, Oxford, UK, July 9–12, 2018, Proceedings 21 (pp. 438–449). Springer International Publishing, 2018
36. Wu L, Cheng JZ, Li S, Lei B, Wang T, Ni D: FUIQA: fetal ultrasound image quality assessment with deep convolutional networks. IEEE transactions on cybernetics. 47(5):1336-49, 2017
37. Kulkarni M, Karande S: Layer-wise training of deep networks using kernel similarity. arXiv preprint http://arxiv.org/abs/1703.07115, 2017
38. Sun Y, Xue B, Zhang M, Yen GG: Completely automated CNN architecture design based on blocks. IEEE transactions on neural networks and learning systems. 31(4):1242-54, 2019
39. Liu Y, Gao Y, Yin W: An improved analysis of stochastic gradient descent with momentum. Advances in Neural Information Processing Systems. 33:18261-71, 2020
40. Bustacara-Medina C, Flórez-Valencia L:An automatic stopping criterion for contrast enhancement using multi-scale top-hat transformation, Sens. Imaging. 20(1):26, 2019
41. Liang J, Zhang M, Liu D, Zeng X, Ojowu O, Zhao K, Li Z, Liu H: Robust ellipse fitting based on sparse combination of data points. IEEE Trans. Image Process. 22(6):2207–2218, 2013
42. Dice LR: Measures of the amount of ecologic association between species. Ecology. 26(3):297–302, 1945
43. Babalola KO, Patenaude B, Aljabar P, Schnabel J, Kennedy D, Crum W, et al: Comparison and Evaluation of Segmentation Techniques for Subcortical Structures in Brain MRI. Med Image ComputComputInterv – MICCAI 2008, Springer, Berlin, Heidelberg, p. 409–16, 2008
44. Shozu K, Komatsu M, Sakai A, Komatsu R, Dozen A, Machino H, Yasutomi S, Arakaki T, Asada K, Kaneko S, Matsuoka R: Model-agnostic method for thoracic wall segmentation in fetal ultrasound videos. Biomolecules. 10(12):1691, 2020
45. Roelfsema NM, Hop WC, Boito SM, Wladimiroff JW: Three-dimensional sonographic measurement of normal fetal brain volume during the second half of pregnancy. American journal of obstetrics and gynecology. 190(1):275-80, 2004

46. Sinclair M, Baumgartner CF, Matthew J, Bai W, Martinez JC, Li Y, Smith S, Knight CL, Kainz B, Hajnal J, King AP: Human-level performance on automatic head biometrics in fetal ultrasound using fully convolutional neural networks. In2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 714–717). IEEE, 2018
47. Al-Bander B, Alzahrani T, Alzahrani S, Williams BM, Zheng Y:Improving fetal head contour detection by object localisation with deep learning, Annual Conference on Medical Image Understanding and Analysis, Springer, pp. 142–150, 2019
48. Liu P, Zhao H, Li P, Cao F: Automated classification and measurement of fetal ultrasound images with attention feature pyramid network, Second Target Recognition and Artificial Intelligence Summit Forum, SPIE, p. 114272R, 2020
49. Li P, Zhao H, Liu P, Cao F: Automated measurement network for accurate segmentation and parameter modification in fetal head ultrasound images. Medical & Biological Engineering & Computing. 58:2879-92, 2020
50. Moccia S, Fiorentino MC, Frontoni E: Mask-R2 CNN: A distance-field regression version of Mask-RCNN for fetal-head delineation in ultrasound images. Int. J. Comput. Assist. Radiol. Surg. 16:1711–1718, 2021
51. Alzubaidi M, Agus M, Shah U, Makhlouf M, Alyafei K, Househ M: Ensemble Transfer Learning for Fetal Head Analysis: From Segmentation to Gestational Age and Weight Prediction. Diagnostics. 12(9):2229, 2022
52. Nie D, Trullo R, Lian J, Wang L, Petitjean C, Ruan S, Wang Q, Shen D: Medical image synthesis with deep convolutional adversarial networks. IEEE Transactions on Biomedical Engineering. 65(12):2720-30, 2018
53. Wang G, Li W, Aertsen M, Deprest J, Ourselin S, Vercauteren T. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. Neurocomputing. 338:34-45, 2019