Review

A review on deep-learning algorithms for fetal ultrasound-image analysis

Maria Chiara Fiorentino [a,*], Francesca Pia Villani [b], Mariachiara Di Cosmo [a],
Emanuele Frontoni [a,c], Sara Moccia [d]

[a] *Department of Information Engineering, Università Politecnica delle Marche, Italy*
[b] *Department of Humanities, Università degli Studi di Macerata, Italy*
[c] *Department of Political Sciences, Communication and International Relations, Università degli Studi di Macerata, Italy*
[d] *The BioRobotics Institute and Department of Excellence in Robotics & AI, Scuola Superiore Sant'Anna, Italy*

ARTICLE INFO

ABSTRACT

Deep-learning (DL) algorithms are becoming the standard for processing ultrasound (US) fetal images. A number of survey papers in the field is today available, but most of them are focusing on a broader area of medical-image analysis or not covering all fetal US DL applications. This paper surveys the most recent work in the field, with a total of 153 research papers published after 2017. Papers are analyzed and commented from both the methodology and the application perspective. We categorized the papers into (i) fetal standard-plane detection, (ii) anatomical structure analysis and (iii) biometry parameter estimation. For each category, main limitations and open issues are presented. Summary tables are included to facilitate the comparison among the different approaches. In addition, emerging applications are also outlined. Publicly-available datasets and performance metrics commonly used to assess algorithm performance are summarized, too. This paper ends with a critical summary of the current state of the art on DL algorithms for fetal US image analysis and a discussion on current challenges that have to be tackled by researchers working in the field to translate the research methodology into actual clinical practice.

1. Introduction

Ultrasound (US) imaging is an imaging modality widely used for the diagnosis, screening and treatment of a large number of diseases, due to its portability, low cost and non-invasive nature (Zaffino et al., 2020). In the years, US imaging has turned out to be the preferred checkup method during pregnancy (Whitworth et al., 2015; Dias et al., 2014). It is commonly used to evaluate fetus's growth and development, as well as to monitor pregnancy and assess clinical suspicion (Bijma et al., 2008). Traditionally, 2-dimensional (2D) US, which produces 2D images that dynamically display fetus's body and internal organs in real-time, is used during pregnancy. More recently, 3-dimensional (3D) US has been made available to support clinicians to better understand fetal spatial anatomic relationships as well as to closely examine suspected fetal anomalies (Huang and Zeng, 2017).

From the clinician's perspective, analyzing fetal US images may be challenging due to the presence of artifacts, such as acoustic shadows, speckle noise, motion blurring and missing boundaries, which are produced as the result of the complex interaction between US waves and mother and fetal biological tissues (Meng et al., 2020a). Further challenges include fast fetus movements, occluded anatomical structures (e.g., due to fetus positioning) and high variability associated with different gestational weeks.

During the last decades, deep learning (DL), and in particular convolutional neural networks (CNNs), have undergone an increasing role in fetal US image analysis to offer decision support to clinicians, and today an extensive literature exists. Survey papers in the field have been published in the last years, even if a number of them focuses on the broader area of medical-image analysis (Liu et al., 2019) or surveys also DL algorithms for US image analysis outside the fetal field (Van Sloun et al., 2019; Akkus et al., 2019; Ouahabi and Taleb-Ahmed, 2021; Shen et al., 2021; Zaffino et al., 2020). Survey papers specifically dealing with US fetal images include: Torrents-Barrena et al. (2019), Song et al. (2021) and Sree and Vasanthanayaki (2019), where segmentation and classification algorithms are covered; Garcia-Canadilla et al. (2020) and Morris and Lopez (2021), that survey methods for fetal cardiology images; Rawat et al. (2018) and Bushra and Shobana (2021), that briefly summarize DL methods for fetal abnormality detection; and Chen et al. (2021c) and Diniz et al. (2020), that analyze research papers from a clinical perspective.

An updated review that surveys the most recent work in the field of fetal US image analysis with DL could be a valuable and compact
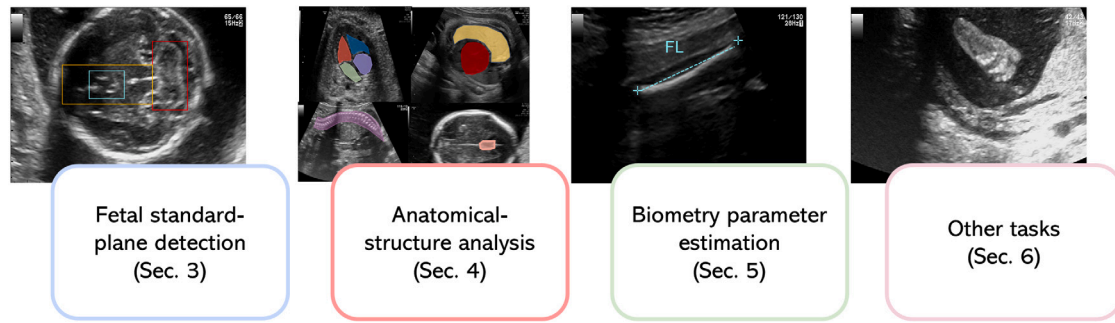

Fig. 1. Summary of the tasks surveyed in this paper.

source of information for young researchers, and a reference overview document for those already working in the field. With this aim, our review starts describing commonly used metrics used for algorithm performance assessment, as well as publicly-available datasets in the field (Section 2.2). As shown in Fig. 1, research papers ranging from fetal standard-plane detection (Section 3) to anatomical structure analysis (Section 4) and fetal-biometry estimation (Section 5) are surveyed. These sections mirror the steps of what is currently done in clinical practice to evaluate fetus well-being. A miscellaneous section (Section 6) is also included, collecting papers on emerging tasks, from less common fetus evaluation applications to probe movement control. For each section, methods are described highlighting pros and cons. Limitations and open issues are further discussed. Summary tables are included to report information on training and testing datasets, as well as gestational age (GA) and achieved performance. A discussion on future directions and open challenges in the field of fetal US analysis with DL concludes this review (Section 7).

1.1. Survey strategy

Our survey strategy started with the following research questions:

- Which are the most investigated tasks addressed using DL in the field of fetal US image analysis?
- Which are the main challenges in regard to fetal examination that are currently tackled by using DL?
- Are the commonly-used datasets sufficient enough for robust DL algorithm development and testing?
- Which are the open issues that still have to be addressed by DL in the field?

With these questions in mind, we outlined a set of keywords for our survey, including: *classification, detection, segmentation, fetal, ultrasound, deep learning* combined together with terms related to fetal examination and organs. The research databases were IEEEXplore, Scopus, Springer-Link, Sciencedirect and Pubmed. For each resulting paper, an extensive review of its reference list was performed. To better focus on the most recent and interesting trends and to not overlap with previous review work, considered the following criteria:

- Papers proposing DL methodologies to process fetal US images
- Indexed papers (Scopus and Web of Science)
- Papers published after 2017
- In regard to journal papers, we only considered papers published in journals ranked (at least) in the 50th percentile, according to CiteScore rank 2020

A summary of the characteristics of the papers analyzed in this review is shown in Fig. 2. Our survey resulted in 153 papers.

Table 1
Contingency table.

		Gold standard	
		Yes	No
Algorithm	Yes	TP	FP
Result	No	FN	TN

2. Metrics and datasets

2.1. Performance metrics

DL algorithms for fetal US image analysis are evaluated using different performance metrics according to the addressed task. Main tasks include classification, which is particularly used for standard plane detection (Section 3) and anatomical structure analysis (Section 4), especially when providing diagnostic support. Besides classification, segmentation and localization/detection are often investigated. The former finds its main application in anatomical structure analysis (Section 4) and biometry estimation (Section 5), as a prior for performing computer-assisted diagnosis (CAD) and/or measuring biometry parameters. Localization and detection are commonly used to localize structures of interest in the image, which finds application in standard plane detection, anatomical structure analysis and biometry estimation.

This section presents the most used quantitative metrics for assessing DL algorithm performance, with the goal to promote fair comparisons. Publicly-available datasets for algorithm development and testing are covered as well. To limit the overlap with the existing literature, we avoid reporting loss functions, as standard loss functions are commonly used in the field of DL for fetal US image analysis. Comprehensive surveys in the field may be found in El Jurdi et al. (2021) and Wang et al. (2022).

For classification tasks, performance is assessed by means of the contingency table (Table 1), with True Positives (TP), True Negatives (TN), False Negatives (FN) and False Positives (FP). Popular classification metrics computed from the contingency table are: (i) Accuracy (Acc) = number of correct predictions ($TP + TN$) divided by the total number of predictions (N); (ii) Recall (Rec) = fraction of actual positives which are correctly identified; (iii) Specificity ($Spec$) = fraction of actual negatives which are correctly identified; (iv) Precision ($Prec$) = proportion of positives which are identified. The F1-score ($F1$) metric, which is equal to the harmonic mean of $Prec$ and Rec, is often used, too. These metrics can be computed at patch, image or patient level.

Other popular metrics are top-1 and top-3 error rates. For top-1 error rate, only the top class (i.e., the one with the highest output probability) is compared with the target label. In the case of top-3 error rate, the target label is compared with the top three predictions (i.e., the three with the highest probability). In both cases, the top score is computed as the number of times the predicted label matched the target label, divided by N, the total number of predictions.
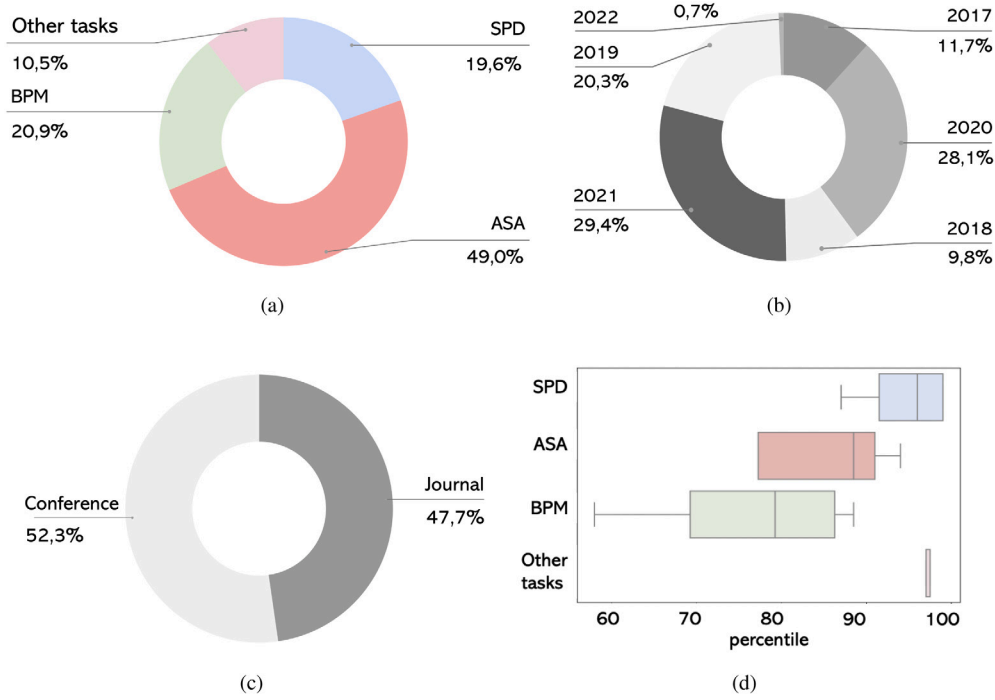
Fig. 2. Summary of the papers included in this work considering: (a) addressed task, (b) year of publication, (c) contribution (journal/conference), (d) percentile according to CiteScore rank 2020 (only for journal papers), SPD = standard-plane detection, ASA = anatomical structure analysis, BPM = biometry parameter estimation.

The Receiver Operating Characteristic (ROC) curve, which shows the performance of a binary classifier as a function of its cut-off threshold, is often reported. The area (AUC) under the ROC curve is used as metric and is interpreted as the probability that the DL model ranks a random positive example more than a random negative example. The higher the AUC (close to 1), the better the model performance.

Even if Acc is widely used, it is not representative when dealing with imbalanced datasets. Therefore, the $F1$ is adopted for imbalanced datasets and multi-class problems. Using the AUC should also be avoided in case of imbalanced datasets, where $Prec - Rec$ curve should be used instead (Sofaer et al., 2019).

As regards segmentation tasks, model performance is commonly evaluated by means of the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU). DSC is equivalent to $F1$ and can be defined as:

$$DSC = \frac{2|X \cap Z|}{|X| + |Z|} = \frac{2TP}{FP + FN + 2TP} = F1 \tag{1}$$

where X and Z are the predicted and the labeled masks, respectively. The IoU is defined as:

$$IoU = \frac{|X \cap Z|}{|X \cup Z|} \tag{2}$$

The Hausdorff distance (HD), which measures how far the labeled segmentation mask is from the predicted segmentation, can be used, too:

$$HD(X, Z) = max(h(X, Z), h(X, Z)) \tag{3}$$

where

$$h(X, Z) = \max_{x \in X} \max_{z \in Z} \|x - z\| \tag{4}$$

DSC and IoU are positively correlated, however IoU tends to penalize single instances of bad segmentation more than DSC, which tends to measure something closer to average performance. Besides DSC and IoU, HD is especially recommended for segmentation tasks with complex boundaries and small thin segments (Aydin et al., 2021), even if it does not consider the progress of curves.

For localization tasks, performance is assessed by means of the Average Precision (AP), which is the $Prec$ averaged across all Rec values between 0 and 1. AP may be seen as the area under the precision–recall curve. The $Prec$ and Rec are commonly computed at various thresholds of IoU between the predicted and labeled bounding box. For detection tasks, where multiple structures have to be detected in the image, mean AP (mAP) is used:

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{5}$$

where K is the number of classes.

For regression tasks, where a numerical value has to be predicted (e.g., for biometry estimation), popular metrics are:

- Mean squared error (MSE), which computes the average squared error between the predicted (y_i) and actual values ($\hat{y}_i$):

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{6}$$

  Due to its differentiable nature, it is often used as loss function during training.
- Mean absolute error (MAE), sometimes referred to as mean absolute deviation, defined as the average of the absolute distance between y_i and $\hat{y}_i$:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{7}$$

- Difference error (DF):

$$DF = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i) \tag{8}$$

- Euclidean distance (ED):

$$ED = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(y_i - \hat{y}_i)^2} \tag{9}$$

MAE is reported along with MSE because it is more robust to outliers. Hence, by relying on the absolute error, MAE avoids errors

Table 2
Performance metrics.

Index	Description
Accuracy (Acc)	$\frac{TP+TN}{N}$
Recall (Rec)	$\frac{TP}{TP+FN}$
Specificity ($Spec$)	$\frac{TN}{TN+FP}$
Precision ($Prec$)	$\frac{TP}{TP+FP}$
Top-1 error rate/top-3 error rate	Top-1 and top-3 errors
AUC	Area Under the Receiver Operating Characteristic curve
AUC by Judd	AUC variant from Judd et al. (2009)
Dice similarity coefficient (DSC)	Eq. (1)
Intersection Over Union (IoU)	Eq. (2)
Hausdorff distance (HD)	Eq. (3)
mean Average precision (mAP)	Eq. (5)
Mean squared error (MSE)	Eq. (6)
Mean absolute error (MAE)	Eq. (7)
Difference (DF)	Eq. (8)
Euclidean distance (ED)	Eq. (9)
Kullback–Leibler divergence (D_{KL})	Eq. (10)
Normalized Scanpath Saliency (NSS)	Eq. (11)

to cancel each other out as when calculating the MSE. Furthermore, MAE does not penalize larger errors more than smaller ones. The advantage of DF and ED over MSE is that, as for MSE, they are measured in the same units as the variable to be regressed.

Recently, researchers are focusing on visual saliency prediction (i.e., predicting human eye fixations on images in the form of saliency maps). Models for saliency prediction can be evaluated using a variety of performance metrics, as described in Bylinskii et al. (2018), which provides an in-depth overview of their advantages and disadvantages. Among them, in the field of US fetal image analysis the Kullback–Leibler divergence (D_{KL}) is often used, which is defined as:

$$D_{KL}(S \parallel F^D) = \sum_{i=1}^{T} F_i^D log(\epsilon + \frac{F_i^D}{\epsilon + S_i}) \qquad (10)$$

where S and F^D are the (predicted) saliency map and (ground truth) continuous fixation map distributions, respectively, ϵ is a regularization constant and T is the total number of fixated pixels.

The Normalized Scanpath Saliency (NSS) (Peters et al., 2005) is used, too. The NSS is computed as:

$$NSS(S, F^B) = \frac{1}{T} \sum_{i=1}^{T} \overline{S}_i F_i^B \qquad (11)$$

where F^B is the ground truth (binary) fixation location map and

$$\overline{S} = \frac{S - \mu(S)}{\sigma(S)} \qquad (12)$$

An AUC variant from Judd et al. (2009) is also used. Here the AUC is built considering TP and FP rates defined as follows. For a given threshold, the TP rate is the ratio of true positives to the total number of fixations, where true positives are saliency map values above threshold at fixated pixels. The FP rate is the ratio of false positives to the total number of saliency map pixels at a given threshold, where false positives are saliency map values above threshold at unfixed pixels.

Table 2 summarizes the performance metrics described in this section.

2.2. Publicly-available datasets

Collecting and sharing high-quality annotated fetal US datasets is not trivial. Labeling large datasets can take a significant amount of time, which may vary according to the task (e.g., pixel-level labeling for segmentation is way more time-consuming than image-level labeling for classification). Data privacy and protection concerns further constitute a barrier to data sharing. To attenuate these issues, international scientific organizations are working to collect and share publicly available databases to encourage algorithm development and fair comparison among algorithms. In the framework of the IEEE International Symposium on Biomedical Imaging (ISBI) and the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), three annotated datasets have been released in the form of a Grand Challenge.[1]

The first challenge in time was the Challenge US: Biometric Measurements from Fetal Ultrasound Images (Rueda et al., 2013), held in conjunction and with the support of ISBI in 2012. The goal of the challenge was to automatically segment fetal abdomen, head, femur and whole fetus for measuring standard biometric parameters, but none attempted the abdomen and whole fetus sub-challenges. Despite the recognized value of the dataset, its size (270 images) does not allow researchers to develop generalizable DL algorithms.

In 2018, with the release of the HC18 challenge dataset[2] (van den Heuvel et al., 2018), the potential of DL for biomerty parameter estimation has been unlocked. The challenge was to develop algorithms to automatically measure fetal head circumference (HC), with 999 and 335 2D US images for training and testing, respectively. Images were annotated by an experienced sonographer.

During ISBI 2021, the A-AFMA ultrasound challenge[3] was organized. The goals were to: (1) detect amniotic fluid and maternal bladder, (2) identify the appropriate landmarks for maximum vertical pocket (MVP) measurement, as to assess amniotic fluid volume.

Recently, a large dataset[4] of routinely acquired maternal-fetal screening US images was made publicly available in Burgos-Artizzu et al. (2020). It consists of 7129 2D training images from 896 patients, categorized into 6 classes: abdomen, brain, femur, thorax, maternal cervix and other. Images were manually labeled by an expert fetal clinician. A test set, which consists of 5271 2D images from 896 patients, was provided. The goal was to boost the research in the field of fetal standard-plane detection.

The publicly-available datasets for the development of DL algorithms for fetal US are summarized in Table 3

3. Fetal standard-plane detection

According to the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) guidelines (Salomon et al., 2019), the use of standardized planes of acquisition improves the reproducibility of (i) fetal biometry assessment and (ii) fetus evaluation. The planes that are typically acquired to extrapolate biometric measurements are fetal abdomen (FASP), brain (FBSP) and femur (FFESP) standard planes. FBSP involves trans-ventricular (FVSP) and trans-thalamic (FTSP) standard planes. Fetus evaluation further requires the acquisition of maternal cervix, fetal heart (including 4 chamber view (4CH), left ventricular outflow tract (LVOT), right ventricular outflow tract (RVOT), three-vessel trachea (3VT), three-vessel view (3VV)), fetal trans-cerebellum standard plane (FCSP), fetal facial standard plane (FFSP) and lumbosacral spine plane (FLVSP). FFSP includes axial (FFASP), coronal and sagittal planes. Visual samples of the most common standard planes are shown in Fig. 3.

In clinical practice, the acquisition of a standard plane is performed manually by clinicians, which move the US probe across the mother's belly until specific anatomical landmarks are visible in the image. Clinical expertise is required to face the high intra-class variability of US standard planes, which is due to, among the others, different gestational weeks, equipment vendors and US-probe angle (Pu et al.,

---

[1] https://grand-challenge.org/challenges/.
[2] https://saras-mesad.grand-challenge.org/.
[3] https://a-afma.grand-challenge.org/.
[4] https://zenodo.org/record/3904280.

Table 3

Publicly available datasets used for fetal ultrasound deep-learning algorithms development. For performance metrics refer to Table 2. MVP: maximum vertical pocket.

Name	Task(s)	Training set size (images/patients)	Testing set size (images/patients)	Annotators	Performance metrics	Gestational age
Challenge US (2012)	Head and femur segmentation & measurement	–/–	284 2D/–	2 and 3, respectively	$Prec$ [%], Acc [%], DSC [%] and distance-based metrics	21, 28 and 33 weeks
HC18 challenge dataset (2018)	Head-circumference estimation	999 2D/[a]	335 2D/[a]	1	ADF [mm], DF [mm], DSC [%], HD [mm]	12 to 35 weeks
A-AFMA ultrasound challenge dataset (2021)	(1) Amniotic fluid and maternal bladder detection (2) MVP detection	–	–	–	mAP [0-1]	–
Burgos-Artizzu et al. (2020)	Standard plane detection of abdomen, brain, heart, femur, maternal cervix and other	7129 2D/896	5271 2D/896	1	Top-1 error rate [%], top-3 error rate [%], Acc [%]	18 to 40 weeks

[a] For the HC18 challenge dataset, only the total number of subjects (551) is reported.

Table 4

Summary of deep-learning algorithms for fetal standard-plane detection (for performance metrics refer to Table 2). FASP = Fetal Abdomen Standard Plane, FFSP = Fetal Facial Standard Plane, FFASP = Fetal Face Axial Standard Plane, 4CH = Four Chamber View, FBSP = Fetal Brain Standard Plane, LVOT = Left Ventricular Outflow Tract, 3VV = Three-Vessel View, RVOT = Right Ventricular Outflow Tract, FFESP = Fetal Femur Standard Plane, FTSP = Fetal Trans-Thalamic Standard Plane, FCSP = Fetal Trans-Cerebellum Standard Plane, FLVSP= Fetal Lumbosacral Spine Standard Plane, FVSP = Fetal Trans-Ventricular Standard Plane. Performance metrics are expressed as the average value over the classes.

Paper	Plane	Training set size	Test set size	Annotators	Performance metrics	Gestational age
Wu et al. (2017a)	FASP	8072 2D (492 videos) (–)	2606 2D (219 videos) (66 subjects)	3	AUC = 0.99, Acc = 0.98, Rec = 0.96, $Spec$ = 0.97	16 to 40 weeks
Yu et al. (2017b)	FFSP	4849 2D[b]	2418 2D[b]	Few	AUC = 0.99, Acc = 0.96, $Prec$ = 0.96, Rec = 0.97, $F1$ = 0.97	20 to 36 weeks
Qu et al. (2020)	FBSPs	15314 2D (155 subjects[a])	3828 2D (155 subjects[a])	Few	Acc = 0.93, $Prec$ = 0.93, Rec = 0.92, $F1$ = 0.93	–
Burgos-Artizzu et al. (2020)	Multiple	7129 2D (896 subjects[a])	5271 2D (896 subjects[a])	1	6.2% top-1 error, 0.27% top-3 error, Acc = 0.94	18 to 40 weeks
Kong et al. (2018)	4CH, FASP, FBSP, FFSPs	17036 2D	5678 2D	1	$Prec$ = 0.98, Rec = 0.98, $F1$ = 0.98	14 to 28 weeks
Liang et al. (2019)	4CH, FASP, FBSP, FFASP, coronal FFSP	17840 2D	4455 2D	–	Acc = 0.99, Rec = 0.96, $Spec$ = 0.99, $F1$ = 0.95	–
Sundaresan et al. (2017)	4CH, LVOT, 3VV	10000 2D (10 subjects)	2178 2D (2 subjects)	1	Error rate = 0.23	20 to 35 weeks
Montero et al. (2021)	FBSP	6498 2D	2249 2D	Few	Acc = 0.81, AUC = 0.86, $F1$ = 0.80	–
He et al. (2021)	Multiple	6974 (920 subjects[a])	2768 (920 subjects[a])	1	Acc = 0.90, mAP = 0.94, $Prec$ = 0.92, Rec = 0.94, $F1$ = 0.93	18 to 28 weeks
Meng et al. (2020c)	4CH, FASP, LVOT, RVOT, FFESP, Lips	12000 2D	5500 2D	Few	$F1$ = 0.77, Rec = 0.77, $Prec$ = 0.78	–
Chen et al. (2017)	FASP, FFASP, 4CH	37376 2D (900 subjects)	13248 2D (331 subjects)	1	Acc = 0.87, $Prec$ = 0.71, Rec = 0.64, $F1$ = 0.64	18 to 40 weeks
Pu et al. (2021a)	FASP, FTSP, FCSP, FLVSP	68296 2D (1199 videos)	16740 2D (244 videos)	Few	Acc = 0.85, $Prec$ = 0.85, Rec = 0.85, $F1$ = 0.85	–
Schlemper et al. (2019)	Multiple	122233 2D	38243 2D	Few	Acc = 0.98, Rec = 0.93, $Prec$ = 0.93, $F1$ = 0.93	18 to 22 weeks
Cai et al. (2018)	FASP	1292 2D (25 subjects)	324 2D (8 subjects)	Few	$Prec$ = 0.96, Rec = 0.96, $F1$ = 0.96	–
Cai et al. (2020)	FASP, FBSP, FFESP	224 videos	56 videos	1	$Prec$ = 0.98, Rec = 0.85, $F1$ = 0.87	–
Lee et al. (2021)	Multiple	1504 2D	752 2D	–	$Prec$ = 0.75, Rec = 0.73, $F1$ = 0.74	Mid-trimester
Dong et al. (2019a)	4CH	5626 2D	1406 2D	Few	mAP = 0.81	14 to 28 weeks
Lin et al. (2018)	FBSP	4800 2D	1153 2D	1	AP = 0.79, Rec = 0.85, $Prec$ = 0.87	14 to 28 weeks
Lin et al. (2019)	FBSP	1451 2D	320 2D	Few	mAP = 0.93	14 to 28 weeks
Zhang et al. (2021)	FASP, FBSP, 4CH	2460 2D	820 2D	2	mAP = 0.95, Acc = 0.95, $Prec$ = 0.95, Rec = 0.93	20 to 34 weeks
Baumgartner et al. (2017)	Multiple	140827 2D and 2438 videos (2694 subjects[a])	109165 2D and 200 videos (2694 subjects[a])	45	$Prec$ = 0.77, Rec = 0.90 $F1$ = 0.80 and IoU = 0.62	18 to 22 weeks

(continued on next page)

Table 4 (*continued*).

| Paper	Plane	Training set size	Test set size	Annotators	Performance metrics	Gestational age
Yaqub et al. (2017)	FVSP	15870 2D	3968 2D	Few	$Acc = 0.95$	Mid-gestation
Gao et al. (2020)	FTSP, FCSP	34586 2D from 441 videos (147 subjects)	60 videos (20 subjects)	2	$mAP = 0.87$	–
Chen et al. (2019a)	Multiple	140827 2D and 2438 videos (2694 subjects[a])	109165 2D and 200 videos (2694 subjects[a])	45	$Prec = 0.89$, $Rec = 0.90$, $F1 = 0.89$	10 to 22 weeks
Tan et al. (2019)	Multiple	22757 2D (2000 subjects[a])	5737 2D (2000 subjects[a])	–	$Acc = 0.70$	–
Dou et al. (2019)	FBSPs	330 3D (330 subjects)	100 3D (100 subjects)	1	$DF = 3.03$ mm, $\theta = 9.36°$	19 to 31 weeks
Yang et al. (2021a)	Multiple	1281 3D	354 3D	4	$DF = 2.31$ mm, $\theta = 10.36°$	–
Yang et al. (2021b)	Uterine standard planes	539 3D (476 subjects[a])	144 3D (476 subjects[a])	4	$DF = 1.82$ mm, $\theta = 7.20°$	19 to 31 weeks
Li et al. (2018)	FVSP, FCSP	50 3D (50 subjects)	22 3D (22 subjects)	1	Mean plane center difference = 3.44 mm, rotation angle = 11.05°	–
Tsai et al. (2021)	Middle sagittal plane	112 3D	28 3D	–	$ED = 0.05$	11 to 13 weeks

[a]Only the total number of subjects is reported.

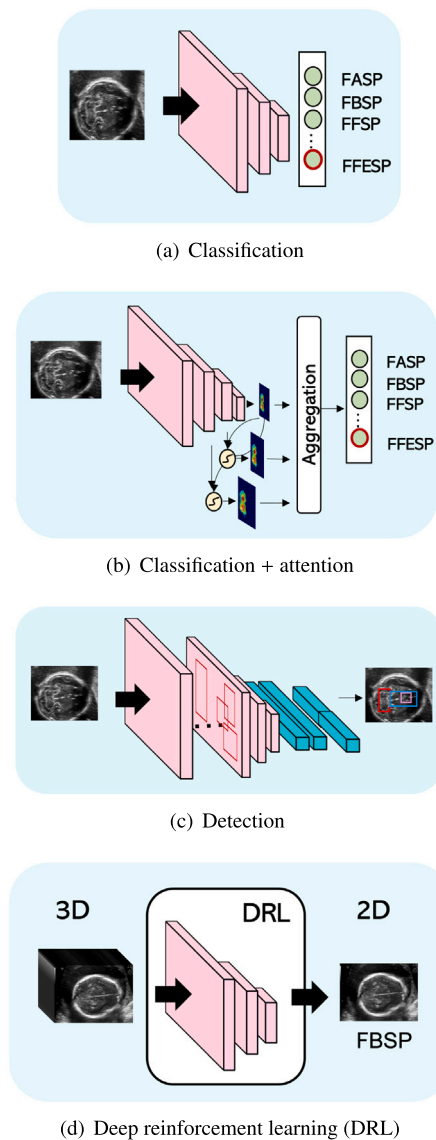[b]The dataset includes unhealthy subjects.
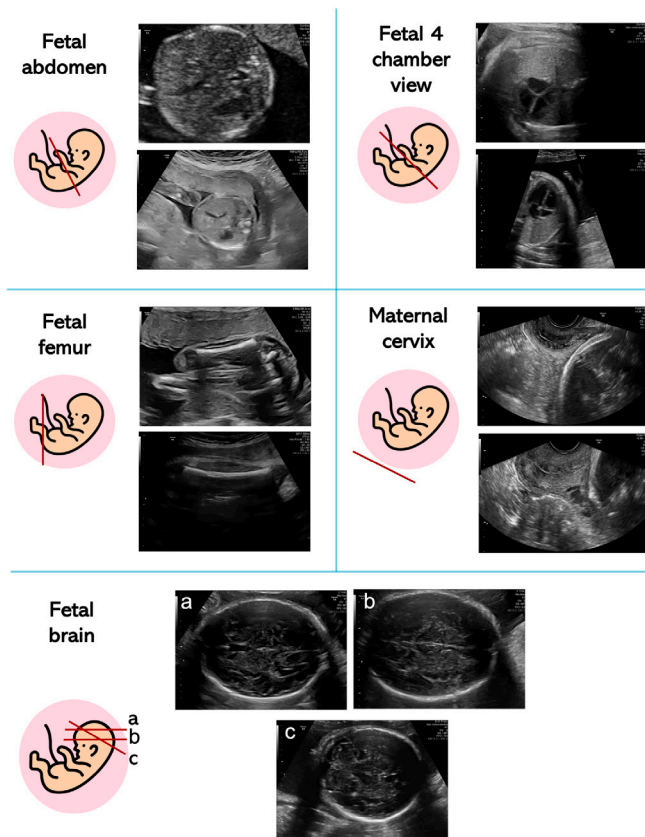
Fig. 3. Visual samples of the most common fetal standard planes.

2021a). Moreover, the anatomical structures that characterize a specific plane could be common to other planes. While approaches in the last decade focused on handcrafted feature-based machine learning (Ni et al., 2014), today DL algorithms are proving to be a valuable tool to tackle these challenges. Table 4 summarizes the DL algorithms for standard-plane detection that are surveyed in this section. An overview of the most common strategies adopted in the literature is shown in Fig. 4.

One of the first DL algorithms for scan plane detection is proposed by Wu et al. (2017a), with the goal of detecting FASP. The algorithm consists of a first CNN that localizes fetal abdomen and a second one that detects the presence in the localized abdomen of stomach bubble and umbilical vein. Despite the promising results, two separate CNNs are needed for localization and classification, respectively, inevitably increasing training and deployment time.

A number of algorithms tackles the problem of fetal standard plane detection as a mere classification problem of 2D still images, following the strategy shown in Fig. 4a. In Yu et al. (2017b), FFSP is detected by fine tuning a shallow classification CNN pre-trained on ImageNet.[5] A shallow classification CNN is also used in Qu et al. (2020) to automatically identify FBSPs. In Burgos-Artizzu et al. (2020), state-of-the-art CNNs are compared to classify 6 planes (FASP, FTSP, FCSP, FVSP, FFESP, and thorax standard planes). The best performing CNN turns out to be DenseNet-169. The authors made the dataset available to foster research in the field (Section 2.2, Table 3). A dense network is also used in Kong et al. (2018) to detect 4CH, FASP, FBSP, FFSPs. Similarly, the work in Liang et al. (2019) proposes an automatic fetal standard plane classification of 4CH, FASP, FBSP, FFASP, and coronal FFSP, based on DenseNet. The network is trained in conjunction with a placenta transferring dataset in order to discover and learn potential relationships between the datasets and to possibly avoid overfitting. In Sundaresan et al. (2017), the detection of heart fetal standard planes (4CH, LVOT, 3VV and not heart) is addressed with a Fully Convolutional Network (FCN), that aims at segmenting the center of fetal heart and classifying the cardiac views in a single step. A generative adversarial network (GAN) is used in Montero et al. (2021) to improve FBSP classification by means of ResNet. In He et al. (2021), a multi-label learning approach is proposed to simultaneously identify multiple standard planes and corresponding anatomical structures. The pipeline is threefold: first a word embedding is used to identify class dependency between standard planes and anatomical structures; then, a graph convolutional network is used to refine the dependency among classes; finally, contrastive learning is used to further increase the discriminative ability. Cross-device classification of 6 anatomical standard planes (4CH, FASP, LVOT, RVOT, FFESP, Lips) is performed in Meng et al. (2020c). An improved feature alignment is used to extract discriminative and domain-invariant features across domains.

---

[5] https://www.image-net.org/.

(a) Classification

(b) Classification + attention

(c) Detection

(d) Deep reinforcement learning (DRL)

Fig. 4. Most common deep-learning strategies to fetal standard-plane detection. FASP = fetal abdomen standard plane, FBSP = fetal brain standard plane, FFSP = fetal facial standard plane, FFESP = fetal femur standard plane.

A couple of research papers extend the classification of standard planes to US video clips. This mimics what is currently done in clinical practice in which the clinician moves the probe until the plane is identified resulting in a 2D data stream. A DL framework is proposed in Chen et al. (2017) to detect FASP, FFASP and 4CH. To process the temporal information encoded in the US videos, the framework makes use of a long short-term memory (LSTM). Similarly, in Pu et al. (2021a), a classification CNN and a recurrent neural network (RNN) are used to detect FASP, FTSP, FCSP, and FLVSP.

To improve standard-plane detection performance and to increase the interpretability of the DL results, attention mechanisms, which endow CNNs to better focus on the most discriminating regions in the US image, have been exploited. The majority of the approaches in this regard implements the strategy shown in Fig. 4b. The work in Schlemper et al. (2019) is among the first works to add a self-gated soft-attention mechanism to a CNN for the detection of 13 US standard planes. Another popular solution in the field is to include sonographer gaze attention mechanism. In Cai et al. (2018), a multi-task framework relying on SonoNet is proposed for FASP detection. The framework is

trained to predict both standard plane and sonographer visual saliency prediction. In Cai et al. (2020), a similar approach is proposed, which further processes US temporal clips via a temporal attention module. The considered planes are the FASP, FBSP and FFESP. Multiple data augmentation methods are exploited in Lee et al. (2021) to improve standard plane classification.

Using classification CNNs may lead to inaccurate detection of standard planes, since it does not involve the detection of any specific anatomical landmark. This, in fact, does not reflect the way clinicians detect standard planes. A different approach investigated in the literature is to accomplish standard-plane detection through anatomical structure detection, as shown in Fig. 4c. This approach is followed in Dong et al. (2019a), where a classification CNN identifies images of the 4CH and a multi-task single-shot detector (SSD) detects the key anatomical structures of the plane as well as the US gain parameter and zoom of the image. In Lin et al. (2018), a Faster R-CNN is used to assess the presence of fetal head specific anatomical structures in US images. A similar approach is followed in Lin et al. (2019). A multi-task framework, consisting of a Faster R-CNN with an additional classification branch, is used to detect 6 key anatomical structures and to evaluate if the head is centered in the image. A multi-task framework is also used in Zhang et al. (2021), in which a CNN inspired by Faster R-CNN is used to both predict the presence of specific anatomical structures in abdomen, brain and heart images and classify these structures. A hybrid approach is proposed in Baumgartner et al. (2017), where a CNN is trained to localize 13 anatomical structures by means of weak supervision provided by image-level labels, thus, without the need for bounding-box annotation during training. In Yaqub et al. (2017), FVSP detection is accomplished by localizing fetal brain by means of a segmentation CNN at first. Then, cavum septum pellucidum (CSP) visibility, fetal brain symmetry, and midline orientation are assessed by means of a number of additional CNNs.

More recently, semi-supervised and self-supervised strategies for standard-plane detection have been investigated. In Gao et al. (2020), a semi-supervised pipeline is proposed to detect FTSP and FCSP from freehand fetal US video. The framework consists of a CNN for feature extraction, a prototypical learning module and a semantic transfer module to automatically label unseen video frames. In Chen et al. (2019a), a self-supervised method for scan plane detection in fetal 2D US images is proposed. Specifically, given an image, two small patches are randomly selected and swapped on the image and this procedure is repeated multiple times. A CNN is trained to restore the altered image back to its original version. The CNN weights are then used to perform classification of several fetal standard planes (FASP, FBSP, FFESP, Kidney, Spine, 4CH, 3VV, RVOT, LVOT and facial profile). A semi-supervised learning approach to classify 13 standard planes is exploited in Tan et al. (2019). Authors' aim is also to test how background class, class imbalance and inter-class similarity influence performances.

A number of researchers is working to perform standard plane detection using deep reinforcement learning (DRL) from 3D fetal US, mainly to mimic what clinicians do and to explore inter-plane dependency. A common DRL scheme is shown in Fig. 4d. The work in Dou et al. (2019) proposes a DRL to localize fetal brain standard planes in US volumes. The DRL framework is equipped with a landmark-aware alignment module that exploits a CNN to detect anatomical landmarks in the US volume. The landmarks are then registered to a plane-specific atlas. The DRL agent's interaction procedure is terminated by means of a RNN module. The approach is further enhanced in Yang et al. (2021a), which designs an adaptive RNN-based termination module to early stop the agent searching. A similar approach is performed in Yang et al. (2021b), which localizes multiple uterine standard planes in 3D simultaneously by multi-agent DRL. The latter is equipped with one-shot neural architecture search (NAS) module. To improve system robustness against the noisy environment, a landmark-aware alignment
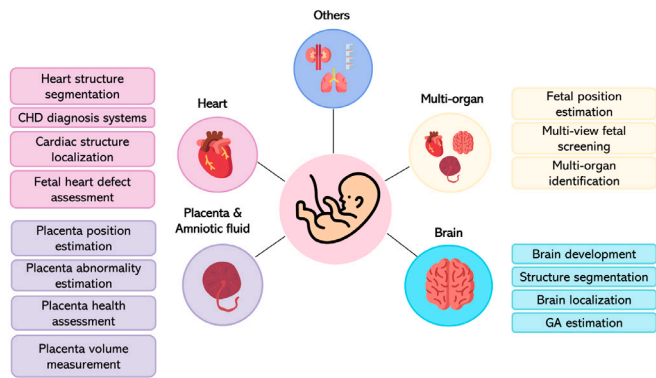
Fig. 5. Overview of the most common tasks in the literature for fetal organs analysis.

model is utilized. The spatial relationship among standard planes is learnt by a RNN.

US volumes are also processed in Li et al. (2018), where 2D image planes are fed to a CNN that predicts the 3D transformation to register each plane to a ground truth standard plane. In Tsai et al. (2021), a GAN is used to estimate the middle sagittal plane from 3D images; however, overfitting may arise considering that (low-dimensional) plane parameters are estimated from high-dimension volumetric data.

3.1. Limitations and open issues for fetal standard-plane detection

Methods used to perform fetal standard plane detection are heterogeneous. Even if classification algorithms have the advantages of being fast and light, they provide only rough localization information of the objects. Attention modules are thus added to better focus on important localization information rather than learning non-useful background information. Detection algorithms, on the contrary, better reflect what is actually performed during a scan plane detection, but with the drawbacks of being heavy and relying on bounding box, which may compromise doctor's understanding of the image.

From the review of the literature on standard-plane detection, a first limitation emerges in the number of considered planes. There are papers (Chen et al., 2017; Yu et al., 2017b; Qu et al., 2020; Pu et al., 2021a) that consider only few planes. In some cases (Cai et al., 2018, 2020; Wu et al., 2017a), only one plane is considered. When focusing on the experimental setup, only few papers refer to US scans on a patient basis. This information is, however, of crucial importance since US scans belonging to a woman should not be part of both training/validation and test sets to avoid bias. A second issue arises in terms of quantitative comparison among the different approaches. The use of different datasets, some of which are really small in size, does not allow making a fair comparison among the solutions proposed in the literature. An important step towards a fair evaluation of methods has been done in Burgos-Artizzu et al. (2020), which released the first dataset in the field. However, only FASP, FBSP, 4CH, maternal cervix, FFESP are considered. In addition, the majority of the surveyed papers neglects the temporal dimension and treats each image individually. This may be considered as an issue for standard-plane detection, as clinicians identify the standard plane by manoeuvring the US probe to the desired scan plane, which results in a 2D data stream. Showing training/validation curves as well as using visual explanation techniques, e.g. Grad-CAM (Selvaraju et al., 2017), should also be considered for a fair assessment of algorithm performance, even in terms of model bias and variance.

4. Anatomical structure analysis

Anatomical structure analysis is one of the most important part of fetus screening, as it provides a direct evidence of possible fetal malformations, as well as placenta localization disorders and premature birth risk. According to ISUOG guidelines (Salomon et al., 2022), from 18–22 weeks of pregnancy organs are already developed. However, organ evaluation during the first trimester (from 11 to 13 + 6 weeks) could also be performed to early detect fetal major anomalies. Fetal screening during mid-trimester involves brain, face, neck, chest/heart, abdomen, skeletal, placenta, umbilical cord, genitalia and cervix.

This section surveys methods for the analysis of anatomical structures, including fetal heart (Section 4.1), brain (Section 4.2), and placenta and amniotic fluid (Section 4.3), for which a relatively rich literature already exists. DL approaches for the analysis of other fetal anatomical structures are grouped in Section 4.4, while Section 4.5 surveys approaches for multi-structure analysis. Fig. 5 summarizes the main tasks addressed in the literature of anatomical structure analysis and Table 5 lists all the papers surveyed in this section.

4.1. Heart

Fetal cardiac evaluation is of crucial importance to detect heart diseases, such as congenital heart diseases (CHDs) and intrauterine growth restriction. Cardiac evaluation generally consists of cardiac-function analysis and heart anatomical evaluation, including heart dimension and shape.

A number of DL approaches in the field focuses on heart and heart-structure detection. The work in Dong et al. (2019b) uses a SSD model with aggregated residual visual blocks to detect anatomical-heart structures such as left atrial pulmonary vein angle, apex cordis, moderator band, and multiple ribs in 4CH. A cardiac-structure localization algorithm is proposed in Patra and Noble (2019). The presence of the heart in the 4CH is detected using a modified VGG-16, then a Faster RCNN model coupled with LSTM layers is used to temporally classify the presence of Foramen ovale, Mitral valve, Tricuspid valve, LV wall and RV wall. In Huang et al. (2017), presence, viewing plane, location and orientation of the fetal heart is predicted by formulating the problem as a multi-task prediction within a recurrent CNN detection framework. In Patra et al. (2017), the same task as in Huang et al. (2017) is formulated as a multi-task learning problem within a hierarchical convolutional model that progressively encodes temporal information throughout the network. In Gao and Noble (2017), spatio-temporal representations of fetal heart are learned by means of an end-to-end two-stream fully CNN for temporal sequence analysis. The goal is to captured motion and appearance features in a weakly supervised manner.

Structure segmentation offers more information than detection, since heart and heart-structure shape is a significant indicator of possible pathologies. Here, encoder–decoder CNNs are often exploited in the literature. In Rachmatullah et al. (2021), a U-Net architecture is used to segment fetal cardiac standard planes to early detect possible heart structural abnormalities. A cascaded network is developed in Xu et al. (2020a) to accurately segment 7 anatomical structures in the 4CH view. The network consists of a dilated sub-network responsible for aggregating both global and local information and two stacked U-Nets. A cascaded U-Net is also used in Xu et al. (2020b) to segment anatomical heart structures. Segmentation of fetal left ventricle from fetal US videos is proposed in Yu et al. (2017a). A dynamic CNN trained with multiscale information and different fine tuning strategies is used. The first cardiac frame is labeled to fine-tune the CNN. As the frames are segmented sequentially, the CNN is fine-tuned dynamically by shallow tuning to fix the latest frame. To differentiate the left ventricle and atrium regions, the mitral valve base points are tracked.

Instance segmentation of heart structures is performed in a number of papers, including (An et al., 2021) where the four cardiac

Table 5

Summary of deep-learning algorithms for anatomical structure analysis (for performance metrics refer to Table 2).

Paper (Year)	Organ	Training set size	Test set size	Annotators	Performance metrics	Gestational age
Dong et al. (2019b)	Heart	–	1991 2D	Few	mAP = 0.93	–
Patra and Noble (2019)	Heart	– (10 subjects)	– (2 subjects)	Few	Acc = 0.82	20 to 35 weeks
Huang et al. (2017)	Heart	90 videos (12 subjects[a])	1 video (12 subjects[a])	Few	–	–
Patra et al. (2017)	Heart	89 videos (10 subjects)	2 videos (2 subjects)	Few	Acc = 0.79	20 to 35 weeks
Gao and Noble (2017)	Heart	350 videos (412 subjects[a])	62 videos (412 subjects[a])	1	ACC = 0.90, $Prec$ = 0.85, Rec = 0.89	≥28 weeks
Rachmatul-lah et al. (2021)	Heart	413 2D (3 subjects[a])[b]	106 2D (3 subjects[a])[b]	Few	IoU = 0.94, Acc = 0.96	18 to 23 weeks
Xu et al. (2020a)	Heart	716 2D (895 subjects[a])	179 2D (895 subjects[a])	Few	DSC = 0.83, Acc = 0.93, AUC = 0.99	–
Xu et al. (2020b)	Heart	1284 2D[b]	428 2D[b]	Few	DSC = 0.85, HD = 3.33, Acc = 0.93	–
Yu et al. (2017a)	Heart	41 videos (40 frames each)	10 videos (40 frames each)	1	HD = 1.26, AD = 0.20, DSC = 0.94	20 to 28 weeks
An et al. (2021)	Heart	512 2D (319 subjects[a])[b]	126 2D (319 subject[a])[b]	10	DSC = 0.78, AP = 0.45	–
Nurmaini et al. (2020)	Heart	624 2D[b]	69 2D[b]	1	Mean DSC = 0.83	18 to 24 weeks
Nurmaini et al. (2021)	Heart	1033 2D (50 subjects[a])	116 2D (50 subjects[a])	1	mAP = 0.96, IoU = 0.79, DSC = 0.90	18 to 24 weeks
Tan et al. (2020)	Heart	– (100 subjects[a])[b]	– (100 subjects[a])[b]	Few	$Prec$ = 0.85, Rec = 0.86, $F1$ = 0.86, AUC = 0.93	–
Dozen et al. (2020)	Heart	410 2D (211 subjects[a])	205 2D (211 subjects[a])	Few	IoU = 0.55, DSC = 0.68	18 to 28 weeks
Komatsu et al. (2021)	Heart	213 videos (363 subjects[a])[b]	34 videos (363 subjects[a])[b]	Few	mAP = 0.70, AUC = 0.83	18 to 34 weeks
Arnaout et al. (2021)	Heart	107823 2D (1326 subjects)[b]	4867591 2D (4666 subjects)[b]	Few	AUC = 0.94, Rec = 0.95, $Spec$ = 0.96	18 to 24 weeks
Qiao et al. (2022)	Heart	2000 2D (2000 subjects)[b]	100 2D (100 subjects)[b]	2	$Prec$ = 0.93, Rec = 0.93	–
Gong et al. (2019)	Heart	3596 2D[b]	400 2D[b]	Few	Acc = 0.85	18 to 39 weeks
Pu et al. (2021b)	Heart	586 videos (minimum 80 and maximum 373 frames) (350 subjects[a])	151 videos (minimum 80 and maximum 373 frames) (350 subjects[a])	Few	Acc = 0.95, Rec = 0.93, $Spec$ = 0.94, $F1$ = 0.95	18 to 36 weeks
Patra and Noble (2020)	Heart	89 videos (39556 2D) (12 subjects)	2 videos (12 subjects)	Few	–	20 to 35 weeks
Wang et al. (2018)	Brain	–	4005 2D (1783 subjects)	Few	DSC = 0.77, IoU = 0.63, HD = 26.40 pixel	28 to 40 weeks
Wu et al. (2020)	Brain	–	448 2D (224 subjects)	3	$Prec$ = 0.79, Rec = 0.74, DSC = 0.77, HD = 0.78	18 to 37 weeks
Singh et al. (2021b)	Brain	588 2D	146 2D	Few	DSC = 0.87, HD = 28.15, Rec = 0.86, $Prec$ = 0.90	18 to 20 weeks
Zhang et al. (2020d)	Brain	718 2D (HC18)	70 2D (HC18)	2	DSC = 0.97, $Prec$ = 0.97, Rec = 0.98, HD = 10.92 mm	12 to 35 weeks
Huang et al. (2018)	Brain	200 3D	45 3D	Few	IoU = 0.63	20 to 29 weeks

Table 5 (*continued*).

Paper (Year)	Organ	Training set size	Test set size	Annotators	Performance metrics	Gestational age
Wyburd et al. (2020)	Brain	271 3D	36 3D	Few	DSC = 0.82, BCE = 0.09	18 to 22 weeks
Venturini et al. (2020)	Brain	480 3D	48 3D	Few	DSC = 0.83, ED = 2.24 mm, HD = 4.51 mm	20 to 25 weeks
Hesse and Namburete (2020)	Brain	138 3D	15 3D	Few	Surface Dice = 0.92, Average Surface Distance = 0.23 mm	20 to 26 weeks
Yang et al. (2020)	Brain	50 3D (100 subjects[a])	50 3D (100 subjects[a])	2	DSC = 0.96, IoU = 0.92, HD = 0.46 mm	20 to 31 weeks
Namburete et al. (2018)	Brain	Not specified 2D (from 599 3D)[b]	Not specified 2D (from 140 3D)[b]	Few	Acc = 0.99, ED = 6.9 mm, IoU = 0.82, HD = 9.3 mm	18 to 34 weeks
Moser et al. (2020)	Brain	885 3D	300 3D	Few	ED = 1.36 mm, HD = 9.05 mm, DSC = 0.94	14 to 31 weeks
Lee et al. (2020)	Brain	8369 2D	869 2D	Few	$RMSE$ = 13.85, MAE = 16.05	≥14 weeks
Namburete et al. (2017)	Brain	326 3D	121 3D	Few	Prediction error = 6.9	18 to 36 weeks
Wyburd et al. (2021)	Brain	689 3D	122 3D	Few	MAE = 4.1, for Sylvian fissure; MAE = 5.1, for Parieto-occipital fissure; MAE = 4.9, for Calcarine fissure	19 to 30 weeks
Xie et al. (2020)	Brain	29419 2D (12780 subjects[a])[b]	4739 2D (12780 subjects[a])[b]	15	mAP = 0.98, Rec = 0.94, DSC = 0.94, Acc = 0.96, $Spec$ = 0.96, AUC = 0.99	22 + 4 weeks (normal)/26 + 3 weeks (abnormal cases)
Hu et al. (2019)	Placenta	1363 2D (247 subjects[a])	205 2D (247 subjects[a])	3	DSC = 0.92, Acc = 0.93	8 to 34 weeks
Hu et al. (2021)	Placenta	10707 2D (321 subjects[a])	2677 2D images (321 subjects[a])	–	Acc = 0.81, Rec = 0.88, $Spec$ = 0.65	11 to 15 weeks
Zimmer et al. (2020)	Placenta	27081 2D (67 subjects[a])	5149 2D (67 subjects[a])	1	DSC = 0.82, HD = 31.84 mm	19 to 33 weeks
Oguz et al. (2018)	Placenta	– (47 subjects[a])	– (47 subjects[a])	2	DSC = 0.86	11 to 15 weeks
Torrents-Barrena et al. (2020)	Placenta	– (70 subjects[a])[b]	– (70 subjects[a])[b]	–	DSC = 0.82	17 to 37 weeks
Looney et al. (2018)	Placenta	1196 3D (3104 subjects[a])	1197 3D (3104 subjects[a])	3	DSC = 0.84, HD = 14.6 mm	11 to 13 weeks
Looney et al. (2017)	Placenta	280 3D (3064 subjects[a])	20 3D (3064 subjects[a])	Few	DSC = 0.73, HD = 27 mm	11 to 15 weeks
Zimmer et al. (2019)	Placenta	30 3D (30 subjects[a])	12 3D (30 subjects[a])	–	DSC = 0.80	24 to 34 weeks
Yang et al. (2017)	Placenta	60 3D (104 subjects[a])	44 3D (104 subjects[a])	10	DSC = 0.64, HD = 24.54 mm	10 to 14 weeks
Schwartz et al. (2021)	Placenta	99 3D (442 subjects[a])[b]	25 3D (442 subjects[a])[b]	–	DSC = 0.88, AUC = 0.89	11 to 14 weeks
Torrents-Barrena et al. (2021)	Placenta	– 3D (60 patients[a])[b]	– 3D (60 patients[a])[b]	–	DSC = 0.76	17 to 37 weeks
Li et al. (2017)	Amniotic Fluid	400 2D	900 2D	Few	Acc = 0.78, IoU = 0.54	22 weeks
Cho et al. (2021)	Amniotic Fluid	310 2D (100 subjects)	125 2D (155 subjects)	1	DSC = 0.87, $Prec$ = 0.89, Rec = 0.87, $Spec$ = 0.99	20 to 36 weeks

Table 5 (*continued*).

Paper (Year)	Organ	Training set size	Test set size	Annotators	Performance metrics	Gestational age
Sun et al. (2021)	Amniotic Fluid	2200 2D (1190 subjects[a])	180 2D (1190 subjects[a])	3	$DSC = 0.86$, $Rec = 0.81$, $Prec = 0.93$, $HD = 15.49$ mm	20 to 36 weeks
Xia et al. (2021)	Lung	6312 2D	701 2D	–	$Acc = 0.83$, $Rec = 0.82$, $Spec = 0.83$, AUC = 0.95	20 to 41 weeks
Chen et al. (2020b)	Lung	206 2D	126 2D	1	$MAE = 1.56$ mm	24 to 41 weeks
Weeras-inghe et al. (2020)	Kidney	40 3D	60 3D	2	$DSC = 0.81$, $IoU = 0.69$, $HD = 8.96$ mm, 20 to 40 weeks	
Franz et al. (2021)	Spine	320 3D	80 3D	–	–	15 to 41 weeks
Chen et al. (2021b)	Spine	−3D (3300 subjects[a])[b]	−3D (3300 subjects[a])[b]	2	$Rec = 0.93$, $Prec = 0.96$, $Acc = 0.94$, $IoU = 0.91$	23 to 31 weeks
Schmidt-Richberg et al. (2017)	Abdomen	126 3D	42 3D	–	$MAE = 2.24$ mm	15 to 38 weeks
Droste et al. (2019)	Abdomen	30 videos	3 videos	Few	AUC by Judd = 0.87, $D_{KL} = 2.16$	–
Wang et al. (2019)	Femur	30 3D (30 subjects)	20 3D (20 subjects)	1	$DSC = 0.91$, $IoU = 0.83$, $HD = 4.08$ mm, $ED = 0.87$ mm	23 to 31 weeks
Cerrolaza et al. (2018)	Skull	52 3D	16 3D	Few	$DSC = 0.83$, $IoU = 0.70$	20 to 36 weeks
Perez-Gonzalez et al. (2020)	Head	10 3D (18 subjects[a])	8 3D (18 subjects[a])	1	$DSC = 0.81$, AUC = 0.82	Second trimester
Wang et al. (2020)	Head	799 2D (HC18)	200 2D (HC18)	Few	$DSC = 0.94$	12 to 35 weeks
Singh et al. (2021a)	Face	54 3D	6 3D	–	$ED = 1.72$ mm	14 to 30 weeks
Chen et al. (2020a)	Face	120 3D	32 3D	–	$IoU = 0.64$	–
Miyagi et al. (2021)	Face	1658 2D	186 2D	3	$Acc = 0.98$, $Rec = 0.93$, $Spec = 0.97$	19 to 38 weeks
Vaze and Namburete (2018)	Adipose tissue	1370 2D	340 2D	3	$Rec = 0.87$, $Spec = 0.99$, $DSC = 0.80$	17 to 41 weeks
Hermawati et al. (2021)	Adipose tissue	65 2D	50 2D	–	$Prec = 0.92$, $Rec = 0.94$, $Spec = 0.95$, $DSC = 0.93$	26 to 41 weeks
Sridar et al. (2019)	Multiple	3109 2D	965 2D	1	$Acc = 0.97$, $Prec = 0.76$, $Rec = 0.75$	18 to 20 weeks
Ishikawa et al. (2019)	Body, Head, Leg	10000 2D	2000 2D	1	$Rec = 0.92\%$	16 to 34 weeks
Sharma et al. (2019)	Multiple	31629 2D	3986 2D	Few	Top-1 $ACC = 0.77$, Top-3 $ACC = 0.94$	–
Alsharid et al. (2020)	Abdomen, Head, Heart, Spine	12808 2D	9979 2D	–	$Prec = 0.96$, $Rec = 0.96$, $DSC = 0.96$	–
Chen et al. (2020c)	Abdomen, Heart, Skull	225825 (free hand) and 30048 (single sweep) 2D	7512 (single sweep) 2D	–	$Acc = 0.73$ and $Acc = 0.89$ for 3- and 4-class classification, respectively	–
Xu et al. (2018)	Kidney, Liver, Spleen	149775 2D	37444 2D	2	$Acc = 0.85$	–
Gao and Noble (2019)	Abdomen, Heart, Skull	365 videos	91 videos	–	$AP = 0.91$	–

Table 5 (*continued*).

Paper (Year)	Organ	Training set size	Test set size	Annotators	Performance metrics	Gestational age
Wu et al. (2017b)	Abdomen, Head	1588 2D	741 2D	1	$DSC = 0.97$, $IoU = 0.96$	19 to 40 weeks
Liu et al. (2021)	Abdomen, Head	4200 2D (Device A)	1516 2D (Device B), 816 2D (Device C)	Few	$DSC = 0.93$, $HD = 12.50$, $IoU = 0.89$, $Rec = 0.92$, $Prec = 0.95$ for abdomen, $DSC = 0.97$, $HD = 6.07$, $IoU = 0.94$, $Rec = 0.97$, $Prec = 0.96$ for head	–
Yang et al. (2019b)	Fetus, Gestational sac, Placenta	60 3D	44 3D	10	$DSC = 0.80$, $HD = 14.13$ mm	10 to 14 weeks

[a]Only the total number of subjects is reported.

[b]The dataset includes unhealthy subjects.

chambers are segmented with a network with three branches: the category branch, the mask branch, and a category-attention branch. This latter is used to correct the instance mis-classification and improve the segmentation accuracy. A Mask-RCNN is used in Nurmaini et al. (2020) to detect and segment the left and right atria, left and right ventricles, aorta and hole. The proposed approach is used to assess fetal heart defects (atrial septal, ventricular septal and atrioventricular septal defects). A Mask RCNN is used in Nurmaini et al. (2021) to segment the fetal heart in 4 different standard views (3VT, 4CH, LVOT, RVOT) and fetal heart chambers in each view to search for possible heart defects.

CHD diagnosis is another popular task investigated in the literature of fetal heart analysis. A hypoplastic left heart syndrome (HLHS) detector is developed in Tan et al. (2020). The detector consists of a SonoNet to detect standard planes (4CH, LVOT, RVOT) and a VGG16 to identify HLHS patients versus healthy subjects. In Dozen et al. (2020), ventricular septal defects are evaluated by means of YOLOv2, which detects the ventricular septum, and U-Net, that is used to segment the cropped ventricular septum area. A calibration module is used to further enhance U-Net output. YOLOv2 is also used in Komatsu et al. (2021) for the detection of abnormalities in cardiac substructures from 4CH and 3VT 2D frames extracted from videos. In Arnaout et al. (2021), an ensemble of deep residual networks are used to classify five heart views and successively classify normal hearts and complex CHD. A modified U-Net is further trained to calculate cardiothoracic parameters such as the cardiothoracic ratio (CTR), cardiac axis (CA) and fractional area change (FAC) for each cardiac chamber. In Qiao et al. (2022), a CNN with a residual learning module linked with guided backpropagation to visualize feature maps is used to diagnose fetal CHD. In Gong et al. (2019), a one-class classification network is proposed to classify patients with CHD and healthy subjects. An GAN is used for data augmentation and compared with state-of-the-art approaches. In Pu et al. (2021b), a YOLOv3 is exploited to detect the four-chamber views at first. Then a mobileNet is used to classify samples into end diastole or end systole. In order to tackle lack of available data in fetal CHD domain, the work in Patra and Noble (2020) proposes an incremental learning approach to build a hierarchical network model that allows for a parallel inclusion of previously unseen anatomical classes without requiring prior data distributions. The goal here is the detection of different anatomical structure in different fetal cardiac views. The pipeline relies on natural hierarchies in US videos and it is built to account for new data in a self-organized fashion.

4.2. Brain

Assessing fetal brain development is crucial to evaluate fetus growth and diagnose brain pathologies. In the clinical practice, several structures are analyzed to assess the fetal brain development and well-being. From the clinician side, this is not effortless due to challenges such as the high intra- and inter-structure variability. Automatic fetal brain analysis commonly includes anatomical structure localization, segmentation, classification, and measurement. Structures considered in the literature include middle cerebral artery, CSP, cerebellum, brainstem, ventricles, thalamus, cortical plates and fissures (Sylvian, Parieto-occipital, calcarine).

Among DL approaches for fetal brain analysis, a number of work focuses on brain-structure segmentation using encoder–decoder architectures. The work in Wang et al. (2018) proposes a DL-based architecture to segment the middle cerebral artery and provide the gate position on Doppler US images to the sonographer. The architecture uses a pre-trained dilated residual network as encoder and dense upsampling convolution blocks as decoder. After the segmentation, the gate position is retrieved as the center of the segmented area. The work in Wu et al. (2020) proposes a deep attention network inspired by the U-Net encoder–decoder architecture to segment the CSP and measure its width to evaluate the presence of anomalies. VGG11 is used as backbone in the encoding path and a channel attention module is introduced between the encoder and the decoder. Similarly, a U-Net inspired CNN (i.e., ResU-Net) is used in Singh et al. (2021b), with the aim to automatically segment the cerebellum. ResU-Net encodes residual blocks and dilated convolutions to recover the spatial resolution lost in the encoder keeping the number of training parameters low. In Zhang et al. (2020d), a U-Net inspired CNN (i.e., MA-Net) is proposed for fetal head circumference segmentation. MA-Net is based on an encoder–decoder architecture consisting of 5 modules: encoder, atrous convolution, pyramid pooling, decoder and residual skip pathway modules.

More recently, a number of segmentation approaches has been exploring 3D architectures for processing the 3D information naturally encoded in 3D US data. A 3D U-Net is used in Huang et al. (2018) to mask out extra-cranial tissues as a preliminary step to simultaneously segment and localize 5 brain structures (lateral ventricles, CSP, thalamus, cerebellum, and cisterna magna). The output of the 3D U-Net is projected in three standard planes (axial, sagittal and coronal), which are then processed by three CNNs, one for each standard plane. Each CNN outputs the 2D segmentation mask of the 5 structures. A 3D U-Net is also used in Wyburd et al. (2020), Venturini et al. (2020), Hesse and Namburete (2020) and Yang et al. (2020). In Wyburd et al. (2020), it is used to segment the fetal cortical plate and measure the depth of the Sylvian fissure on a dataset annotated by expert clinicians using an atlas Similarly, in Venturini et al. (2020) a multi-task CNN based on 3D U-Net is used for the automatic segmentation of the white matter, thalamus, brainstem, and cerebellum. In Hesse and Namburete (2020), the authors propose a 3D U-Net combined with active contours for CSP segmentation. In Yang et al. (2020), a 3D U-Net is trained to segment the whole fetal head. The 3D U-Net is combined with a hybrid attention scheme to enhance the feature maps.

Brain-structure segmentation is further explored and included in a multi task context in Namburete et al. (2018). The authors propose an approach to align 3D US brain volumes, recovered from a stack of axial slices, to a coordinate system based on skull boundaries, eye socket location, and head pose. A multi-task FCN is trained to address the problem of 3D fetal brain localization, structural segmentation, and alignment. A method for brain localization is also proposed in Moser

et al. (2020), which explores the use of an end-to-end 3D CNN for automated brain localization and extraction from 3D fetal US. Differently from Namburete et al. (2018), which predicts the position of the brain from 2D slices extracted from 3D volumes, this is a fully 3D approach and relies on a modified 3D U-Net for brain extraction.

Other than structure segmentation and localization, researchers are also working on GA estimation and brain development analysis. An approach to GA estimation from 2D trans-thalamic US images is proposed in Lee et al. (2020). The authors propose a Bayesian Neural Network (BNN) with a VGG-16 backbone and an auxiliary regression model, which are trained to predict calibrated aleatoric and epistemic uncertainties on GA. Methods to evaluate brain development are investigated in Namburete et al. (2017) and Wyburd et al. (2021). The work in Namburete et al. (2017) presents a model to predict brain maturation from 3D US through the use of a 3D Convolutional Regression Network. In Wyburd et al. (2021), the authors propose a method to estimate the development of three brain fissures (i.e., Sylvian, Parieto-occipital and Calcarine) from 3D US volumes, by predicting the fetal GA based on their respective morphology. The regions relative to each fissure are extracted from the US volumes and passed to three separate ResNet to predict the GA of each region.

As for CAD of brain pathology, in Xie et al. (2020) the authors propose a DL-based pipeline to support the diagnosis of brain lesion from 2D images. The pipeline relies on three architectures for (i) cranio-cerebral segmentation, using U-Net with dilated convolutional, (ii) image classification in normal/abnormal, using VGG-19 pre-trained on ImageNet, and (iii) lesion localization, using Grad-CAM. The lesion localization, in the form of Grad-CAM activation heatmaps, is visually evaluated by a doctor.

4.3. Placenta & amniotic fluid

The placenta is an organ that provides oxygen and nutrients to the fetus, ensures thermo-regulation, and removes waste products from the fetus's blood. This structure is closely related to fetus health: abnormal placental function may affect the development of the fetus, and in severe cases, even endanger the life of the fetus. Given its relevance, several clinical assessments are performed to evaluate the placenta condition along with the fetus health. In this scenario, placenta segmentation may provide automatic quantification of placenta volume and morphology.

A number of approaches performs placenta segmentation from 2D US images. The approach proposed in Hu et al. (2019) uses a U-Net-inspired CNN. The CNN is modified adding a layer to detect acoustic shadow and improve the segmentation accuracy. Similarly, in Hu et al. (2021) a U-Net-based CNN is used for placenta segmentation. An EfficientNet is then used for the classification of normal or abnormal placenta. A U-Net-like architecture for placenta segmentation is also used in Zimmer et al. (2020). The authors introduce an auxiliary classification task, incorporating the prediction of the placental position (anterior or posterior) in the U-Net architecture and improve the segmentation accuracy. In Oguz et al. (2018), three 2D cross-sectional images manually extracted from 3D US, are used to train an encoder–decoder CNN architecture. An atlas-based joint label fusion algorithm is then applied to the CNN output to combine the three predictions and enhance segmentation performance. Predictions from CNN and joint label fusion are combined via a random forest model to obtain the final segmentation. In Torrents-Barrena et al. (2020), stacked GANs are used to automatically segment placenta along with its peripheral vasculature, as well as acoustic shadows, with the goal to provide surgeons with context awareness during twin-to-twin transfusion syndrome (TTTS) treatment.

Other researchers are working on placenta segmentation from 3D US data. The work in Looney et al. (2018) uses a 3D U-Net. For the same goal, in Looney et al. (2017) the authors use an open-source CNN trained with labels obtained through a semi-automatic Random

Walker strategy. The work in Zimmer et al. (2019) proposes a three stage pipeline for whole placenta segmentation and volume estimation from multi-view 3D US volumes. The pipeline includes: (i) multi-view acquisition through a multiplexer device that fixes 3 probes in an angle of 30 degrees to each other and switches the view from one probe to the other almost in real time; (ii) voxel-wise fusion to combine the multiple views; (iii) segmentation based on a 3D U-Net. The work in Yang et al. (2017) proposes an approach to automatically segment placenta, fetus and gestational sac. The first step includes a customized 3D CNN with long skip connections to segment the 3 structures. A RNN is used to add contextual knowledge and refine the semantic segmentation. Finally, a hierarchical deep supervision mechanism is used to boost the flow of information in the RNN and improve the segmentation performance.

Volumes of 3D US obtained from singleton pregnancies were automatically segmented by Schwartz et al. (2021): their CNN pipeline combines two 2D and one 3D models with downsampling/upsampling architectures. The placental volumes, derived from the automatic segmentation model, are used to train multivariable logistic-regression classifiers to predict small-for-gestational-age infants born under the 10th/5th centiles (SGA10/SGA5).

An interesting approach to placenta segmentation from multimodal imaging (i.e., 3D US and magnetic resonance imaging) is proposed in Torrents-Barrena et al. (2021). The approach combines radiomics and DL to efficiently segment intra-uterine tissues.

Amniotic fluid plays an important role in fetal well-being and development. Amniotic fluid has a myriad of functions: it protects fetus and umbilical cord, prevents infections and provides the necessary growth factors to allow normal development and growth of fetal organs. In the clinical practice, to assess the sufficiency of amniotic fluid quantity, the amniotic fluid index is used. This index is calculated dividing the maternal abdomen into 4 quadrants using the midline and the umbilicus, then the deepest pocket of amniotic fluid is evaluated in each quadrant. The index is given by the sum of the 4 measurements. With a view to automatize the index computation, amniotic fluid segmentation is a crucial task.

In Li et al. (2017), an encoder–decoder architecture with VGG16 as backbone is used to segment amniotic fluid and fetal body in 2D US images. In Cho et al. (2021), a 2-step framework is used to segment the amniotic fluid pocket and measure the amniotic fluid index. The segmentation step is performed with a modified U-Net architecture, called AF-net, which combines atrous convolution and multi-scale side-input and side-output layers. This approach is further improved in Sun et al. (2021), which proposes a dual-path network to segment the amniotic fluid volume from 2D US images. The primary path consists of AF-net, while the secondary path is an auxiliary CNN used to remove reverberation artifacts and complement the primary path prediction. The final segmentation output is obtained combining results of primary and secondary paths.

4.4. Others

4.4.1. Lungs
Immaturity of fetal lung development is the primarily cause of neonatal respiratory morbidity. Quantitative US imaging is often used as a non-invasive tool for fetal lung maturity assessment, through US visualization at 4CH level. The normal fetal lung has very similar echogenicity to the adjacent liver, but with slightly different texture; and in presence of abnormalities either increasing or decreasing in echogenicity can be seen. Recently, DL approaches have been introduced to tackle the challenges of lung analysis from fetal US image, especially for estimating GA.

The framework in Xia et al. (2021) is based on DenseNet, which is developed for processing US images of 4CH plane and to predict the GAs based on 3 classes: class I (from 20 to 29 + 6 weeks), class II (from 30 to 36 + 6 weeks) and class III (from 37 to 41 + 6 weeks). For the same aim, the work in Chen et al. (2020b) proposes a two-stage transfer

learning approach. A U-Net-like architecture with residual connection is first trained to recognize samples of fetal lung regions from other regions taken from the US images; then, the pre-trained U-Net is tuned on fetal lung region samples only to regress the corresponding GA, and thus its maturation degree. The study is performed on 4CH US scans, manually segmenting the lung region. However, this work provides only an indirect prediction of the lung maturation degree, as it relies on the direct correlation between the lung maturity and the GA, which does not always exist.

4.4.2. Kidney

In-utero assessment of kidney is crucial to perform early diagnosis of renal pathologies. Poor kidney development is known to be associated with increased risk of kidney disease into adulthood. At the same time, abnormal fetal growth is associated to a reduced functionality of kidneys after birth (Luyckx and Brenner, 2015; Senra et al., 2020). In this field, the work in Weerasinghe et al. (2020) proposes a 3D U-Net to perform kidney segmentation from 3D B-mode and Power Doppler volumes. Despite the small number of testing volumes, this is an interesting study that may pave the way for the development of early diagnosis tools for fetal kidney.

4.4.3. Spine

Fetal spine length provides insights into the fetal growth as it is affected by a variety of malformations (spina bifida, meningocele, diastematomyelia, vertebral segmentation anomalies, sacral agenesis, spinal dysgenesis, spondylothoracic or spondylocostal dysplasia). To provide an assessment of fetal spine, the work in Franz et al. (2021) focuses on the identification of spine centerline from 3D US scans. Spine segmentation is performed with a CNN that processes images at multiple scales and different fields of view. The prediction is used as input for a model-based tracing algorithm responsible to draw spine centerline. For a similar aim, the work in Chen et al. (2021b) presents an approach to identify the spina-bifida and segment it. The approach is based on an efficient implementation of U-Net replacing standard convolution with the octave convolution proposed in Chen et al. (2019b).

This approach achieves high recognition accuracy, good segmentation accuracy, and short running time, but it is limited on few cases, without dividing among different spina bifida types and with an imbalanced presence on malformation cases compared to healthy ones.

4.4.4. Abdomen

Fetal abdomen assessment is performed to evaluate important prognostic parameters of neonatal morbidity and mortality and to assess fetal growth. Segmentation of the abdominal outline is of interest for measuring fetal abdominal circumference. For this purpose, the work in Schmidt-Richberg et al. (2017) develops an approach based on CNNs to extract image features to be integrated in a deformable model. Fetal abdomen localization is achieved in Droste et al. (2019) using an encoder–decoder architecture for US video saliency prediction. The encoder part consists of a truncated SonoNet, while the spatio-temporal decoder of the network is made up of a bidirectional gated-recurrent-unit recurrent convolutional network (GRU-RCN).

4.4.5. Femur

Volume and length of fetal femur have unique importance in fetal weight estimation. However, it is hard to be evaluated through US, due to the difficulty in locating the tips of the femur, boundary deficiency and ambiguity for tissues' low contrasts, and variation of pose, shape and size of this structure. In Wang et al. (2019), the authors develop a unified framework for simultaneous segmentation and landmark localization of fetal femur in prenatal US volumes: fetal femur ROI is first identified through a U-Net model, then segmentation and landmark localization branches receive the common features of ROI extracted by the shared layers and generate task-specific descriptors.

4.4.6. Skull and head

The evaluation of fetal head is a critical part of sonographic examination. This evaluation is dependent on operator experience and US image intrinsic characteristics. To attenuate these issues, DL algorithms in the literature mainly focus on the segmentation of the fetal head. The work in Cerrolaza et al. (2018) proposes a two-stage approach for skull segmentation in fetal 3D US. A 3D U-Net is used to roughly segment the skull. This segmentation concatenated with two additional channels (i.e., the US-wave incidence angle map and the US shadow casting map) and fed to a second 3D U-Net. The work in Perez-Gonzalez et al. (2020) aims at merging several partially occluded US volumes, acquired by placing the US transducer at different projections of the fetal head, to compound a new US volume containing the whole brain anatomy. For this aim, the authors propose a pipeline of 4 CNNs. The first 2 CNNs follow what is done in Perez-Gonzalez et al. (2020), while the additional 2 CNNs perform US-volume registration. A different approach is proposed in Wang et al. (2020), with the aim of segmenting fetal head. A CycleGAN is trained with a set of unpaired images and auxiliary masks obtained from a shape prior model, to generate the pseudo-labels corresponding to each of the training images. The CycleGAN is equipped with a Variational Auto-encoder based discriminator and a Discriminator-guided Generator Channel Calibration module which calibrates the pseudo-label generator using the discriminator's feedback to improve the pseudo-labels.

4.4.7. Face

Accurate detection and visualization of fetal face position and orientation is crucial in prenatal diagnosis, growth monitoring and detection of fetal anomalies. A U-Net inspired encoder–decoder 3D CNN is proposed in Singh et al. (2021a) to segment fetal face from 3D US volumes. In Chen et al. (2020a), an RPN-based object detection framework is proposed to detect landmarks in 3D facial US volumes. Predictions from the RPN architecture are further refined with a distance-based graph prior to produce the final bounding box for each landmark. Fetal facial expression is another important aspect to take into consideration for the evaluation of the fetal brain and central nervous system. In Miyagi et al. (2021), a custom-built CNN classifier is trained to recognize 7 fetal expressions in 2D US images.

4.4.8. Adipose tissue

Observation and evaluation on fetal adipose tissue is significant to determine the growth and nutritional adequacy of the fetus. The work in Vaze and Namburete (2018) performs fetal adipose tissue segmentation through a U-Net with depth-wise separable convolutions.

The work in Hermawati et al. (2021) proposes a framework for detecting and segmenting the fetal thigh cross-sectional area of adipose tissue. The framework relies on Faster R-CNN for localizing the regions containing the fetal thigh cross-section, on which threshold-based segmentation is performed to allow measurement of adipose tissue thickness.

4.5. Multi-organ analysis

DL researchers are working on multi-organ analysis to reproduce the actual clinical screening, which accounts for a comprehensive list of anatomical structures.

A number of DL learning algorithms in the field focuses on classification tasks. The work in Sridar et al. (2019) proposes a method to automatically classify 14 different fetal structures (abdomen, arm, blood vessels, cord insertion, face, femur and humerus, foot, genitals, head, heart, kidney, leg, spine, hand) from US images by fusing information from both cropped regions of fetal structures and whole images. Two CNNs pre-trained on ImageNet are used as feature extractors. The features are classified by means of support vector machines. Similarly, the work in Ishikawa et al. (2019) fine-tunes a VGG16 to classify US images in four classes (head, body, leg and other). This is needed as

Fig. 6. Most addressed tasks (from left to right) in the field of anatomical structure analysis. Lilac = heart, cyan = brain, purple = placenta and amniotic fluid, yellow = multi-organ, blue = others, temp = temporal information.

preliminary step towards the estimation of fetal position. Based on the output of Grad-CAM, the body parts position in the images is estimated. Spatio-temporal classification is performed in Sharma et al. (2019), in which a LSTM is used to classify 11 anatomical categories (heart, background, brain with skull and neck, Doppler maternal anatomy, spine, abdomen, nose and lips, kidneys, face side profile, femur, other).

In Alsharid et al. (2020), the authors propose a method for multi-organ (abdomen, head, heart, spine) classification from the integration of fetal US images with corresponding textual descriptions. The authors exploit a curriculum learning approach to train a NLP-based fetal US image captioning model with a dataset prepared using real-world US videos along with synchronized and transcribed sonographer speech recordings: 12 808 and 9979 image-caption pairs are used for training and testing sessions, respectively.

An unsupervised approach to multi-organ classification is proposed in Chen et al. (2020c), which develops a cross-device and cross-anatomy adaptation network to classify heart, abdomen and skull of an unlabeled single-sweep video dataset guided by knowledge of a labeled free-hand scanning protocol video dataset. The network consists of encoder, projection layer, anatomy classifier, domain classifier and two mutual information discriminators.

A different approach is used in Xu et al. (2018), where a multi-task learning framework is proposed to classify 11 different views related to abdominal organs and detect 14 landmarks from kidney, liver and spleen. This framework relies on a shared ResNet encoder and two branches, for classification and landmark detection.

Multi-organ analysis also involves DL algorithms for localization and segmentation tasks. In a weakly supervised fashion, the work in Gao and Noble (2019) proposes a model consisting in a CNN for extracting features and an attention-gated LSTM to localize skull, abdomen and heart in consecutive frames, considering also non-standard planes. A pipeline to segment fetal head and abdomen is proposed in Wu et al. (2017b). The pipeline consists of three cascaded FCN and each FCN take as input the summation of fetal US image with prediction map from the previous level to refine boundaries of the anatomy segmentation. In Liu et al. (2021), segmentation of fetal head and abdomen is performed based on style transfer, a technique used to extract the semantic content from a target image and render it to a source image. The authors use this technique to tackle the problem of appearance shift: a U-Net is combined with style transfer based on dynamic instance normalization to perform organ segmentation from US images acquired with three different devices.

In Yang et al. (2019b), a framework to simultaneously segment fetus, gestational sac, and placenta is proposed. The framework consists of a 3D CNN, which explores spatial intensity concurrency, and a RNN, which encodes spatial sequential to improve boundary refinement.
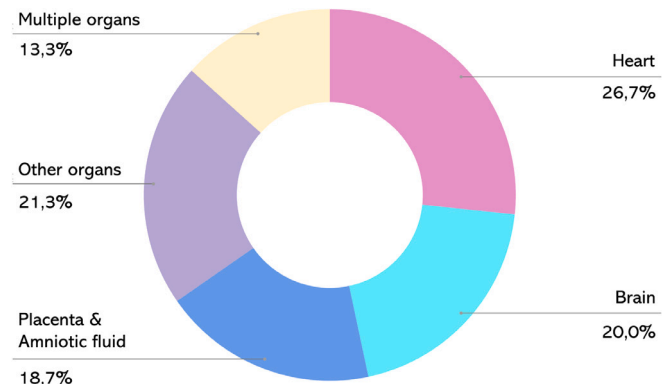
Fig. 7. Number of papers surveyed in Section 4 according to the anatomical structure.

4.6. Limitations and open issues for anatomical structure analysis

As shown in Fig. 6, segmentation is the most addressed task for all the anatomical structures. However, most of the proposed methodologies consist of 2D U-Net-based architectures, while more advanced architectures exploited in closer fields (e.g. adversarial segmentation, spatio-temporal processing) are less investigated. Interestingly, priors relevant to anatomical structure shape have not been fully exploited yet. However, shape-constraint DL strategies have been recently proposed in the literature to drive the output of segmentation CNN towards the desired structure shape, for example by using adversarial learning (Bohlender et al., 2021; Casella et al., 2021).

Up to now, the most studied anatomical structures are heart and brain, contributing with 26.7% and 20.0% of the surveyed papers, respectively, as shown in Fig. 7.

As regards fetal heart analysis, papers in literature deal with the most common screening examinations such as 4CH and LVOT/RVOT views assessment, ventricular and atrial chambers evaluation and fetal heart location and orientation estimation along with fetal heart motion tracking. This reflects what is currently considered by ISUOG guidelines to be the best fetal heart evaluation practice (Salomon et al., 2022). However, at present such assessments are performed as distinct tasks, trained and evaluated on different datasets. The heterogeneity of these approaches does not allow comparison among algorithms: this is particular evident in CHD assessment, which is evaluated considering different approaches (from cardiac structure segmentation to global image classification).

In addition, for a better functional and anatomical assessment of the fetal heart, volume sonography (3D/4D spatiotemporal image correlation, STIC) and tissue Doppler can also be used, as declared in the ISUOG guidelines (Salomon et al., 2022). However, these techniques

Table 6

Summary of deep-learning algorithms for biometry parameter estimation (for performance metrics refer to Table 2). HC: Head circumference, BPD: Biparietal diameter, OFD: occipito-frontal diameter, AC: Abdominal circumference, FL: Femur length, TCD: trans-cerebellar diameter.

Paper	Biometry	Training set size	Test set size	Annotators	Performance metrics	Gestational age
Sinclair et al. (2018)	HC, BPD	2164 2D	100 2D	2	$DSC = 0.98$, $MAE = 1.3$ mm	18 to 22 weeks
Rong et al. (2019)	HC	999 2D (HC18)	335 2D (HC18)	1	$DSC = 0.95$, $MAE = 2.44$ mm	12 to 35 weeks
Skeika et al. (2020)	HC	999 2D (HC18)	297 2D (HC18)	1	$DSC = 0.97$, $MAE = 1.89$ mm	12 to 35 weeks
Zeng et al. (2021)	HC	999 2D (HC18)	335 2D (HC18)	1	$DSC = 0.98$, $MAE = 1.77$ mm	12 to 35 weeks
Aji et al. (2019)	HC	999 2D (HC18)	335 2D (HC18)	Few	$DSC = 0.76$, average error = 14.96%	12 to 35 weeks
Qiao and Zulkernine (2020)	HC	849 2D (HC18)	150 2D (HC18)	1	$DSC = 0.97$, $MAE = 2.27$ mm	12 to 35 weeks
Oghli et al. (2020)	HC	HC18	HC18	1	$DSC = 0.95$, $HD = 4.5$ mm	12 to 35 weeks
Bhalla et al. (2021)	HC	HC18	HC18	1	$MAE = 2.16$ mm	12 to 35 weeks
Sobhaninia et al. (2019)	HC	HC18	HC18	1	$DSC = 0.97$, $MAE = 2.12$ mm	12 to 35 weeks
Sobhaninia et al. (2020)	HC	849 2D (HC18)	150 2D (HC18)	1	$DSC = 0.92$, $MAE = 2.22$ mm	12 to 35 weeks
Li et al. (2020)	HC, BPD, OFD	999 2D (HC18)	335 2D (HC18)	1	$DSC = 0.97$, $MAE = 1.81$ mm	12 to 35 weeks
Kim et al. (2019)	HC, BPD	102 2D	70 2D	2	$ACC = 0.87$, success rate = 0.93	18 to 35 weeks
Al-Bander et al. (2019)	HC	999 2D (HC18)	335 2D (HC18)	1	$DSC = 0.97$, $MAE = 2.33$ mm	12 to 35 weeks
Fiorentino et al. (2021)	HC	999 2D (HC18)	335 2D (HC18)	1	$DSC = 0.97$, $MAE = 1.90$ mm	12 to 35 weeks
Moccia et al. (2021)	HC	999 2D (HC18)	335 2D (HC18)	1	$DSC = 0.98$, $MAE = 1.95$ mm	12 to 35 weeks
Meng et al. (2020d)	HC	905 2D (HC18)	94 2D (HC18)	1	$DSC = 0.97$	12 to 35 weeks
Budd et al. (2019)	HC	2848 2D (2000 subjects[a])	540 2D (2000 subjects[a])	45	$DSC = 0.98$, $MAE = 1.81$ mm	12 to 22 weeks
Zhang et al. (2020a)	HC	800 2D (HC18)	199 2D (HC18)	1	$MAE = 4.52$ mm	12 to 35 weeks
Zhang et al. (2020b)	HC	800 2D (HC18)	199 2D (HC18)	1	$MAE = 4.78$ mm	12 to 35 weeks
Jang et al. (2017)	AC	56 2D (1 subject[a])	32 2D (1 subject[a])	2	$DSC = 0.85$, $ACC = 0.79$	20 to 34 weeks
Kim et al. (2018)	AC	– (77 subjects[a])	– (77 subjects[a])	–	$DSC = 0.92$, $ACC = 0.87$	–
Cengiz and Yaqub (2021)	CRL	697 2D	545 2D	1	$DSC = 0.93$, $IoU = 0.88$, $MAE = 3.87$ mm	11 to 15 weeks
Chen et al. (2020d)	LV	4379 2D[b]	500 2D[b]	3	$Acc = 0.96$, $Prec = 0.98$, $Spec = 0.90$, $MAE = 1.80$ mm	–
Zhu et al. (2021)	FL	2300 2D (435 subjects[a])	310 2D (435 subjects[a])	1	$DSC = 0.92$, $MAE = 0.46$ mm	–
Chen et al. (2021a)	CTR	1669 2D	417 2D	2	$DSC = 0.93$	18 to 39 weeks
Bano et al. (2021)	HC, AC, FL	262 2D (42 subjects[a])	87 2D (42 subjects[a])	1	$MAE = 2.67$ mm (HC), $MAE = 3.77$ mm (AC), $MAE = 2.10$ mm (FL)	–
Płotka et al. (2021)	HC, AC, FL	274275 2D (560 patients)	57001 2D (140 patients)	6	$DSC = 0.96$, $MAE = 2.9$ mm (HC), $MAE = 3.8$ mm (AC), $MAE = 0.8$ mm (FL)	15 to 38 weeks
Ryou et al. (2019)	CRL, HC, AC	44 3D	21 3D	Few	$IoU = 0.92$ for abdomen and head, $IoU = 0.66$ for limbs segmentation, $MAE = 2.24$ mm	11 to 14 weeks
Oghli et al. (2021)	AC, BPD, FL, HC	999 2D (HC18) + 1154 2D (551 subjects[a])	335 2D (HC18) + 180 2D (551 subjects[a])	2	$DSC = 0.98$, $HD = 1.14$ mm, average perpendicular distance = 0.2 mm	12 to 35 weeks
Gao et al. (2021)	HC, TCD	937 2D (937 subjects)	913 2D (913 subjects)	1	$MAE = 2.04$ mm	–
Prieto et al. (2021)	BPD, HC, CRL, AC, FL	147855 2D (4433 subjects)	7233 2D (2491 subjects)	Few	$Acc = 0.93$, $IoU = 0.91$, $MAE = 1.89$ cm	13 to 18 weeks
Rasheed et al. (2021)	HC, BPD	30000 2D	1000 videos	1	$Acc = 0.96$	18 to 42 weeks

[a] Only the total number of subjects is reported.
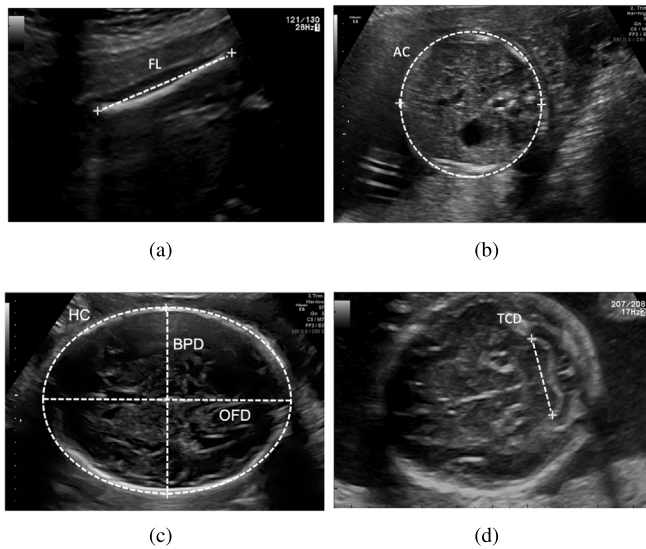[b] The dataset includes unhealthy subjects.

Fig. 8. Common biometry parameters. Transverse section of: (a) femur length (FL), (b) abdominal circumference (AC), (c) head circumference (HC), biparietal diameter (BPD) and occipito-frontal diameter (OFD), (d) trans-cerebellar diameter (TCD).

(a) Segmentation (along with possible head localization pre-processing) + Ellipse fitting

(b) Multi-task framework

(c) Region proposal along with boundary regression + Ellipse fitting

(d) Direct regression

Fig. 9. Most common deep-learning strategies to fetal HC measurement.

have not been exploited yet in the domain of DL for fetal image analysis.

Applications for DL in fetal brain anatomical structure analysis are heterogeneous as well and each one focuses on a particular task, such as brain localization, segmentation, classification or GA estimation. Given the fragmented nature of the literature available in this field, comparing the different approaches is challenging. The lack of public datasets further hampers a fair comparison among the different approaches.

Despite the variety of structures that are evaluated by clinicians during fetal screening, only part of these structures has been considered in the literature so far (Fig. 7). More importantly, most of the work does not focus on detecting problematic pregnancies, which is the main goal of fetal screening, but is limited to anatomical structure detection or segmentation. Some groups are working in this direction (e.g., Schwartz et al. (2021), Arnaout et al. (2021) and Xie et al. (2020) for detecting small for GA infants, CHD, and abnormal brain, respectively), but there are still open research questions.

The same considerations can be applied to fetal placenta evaluation domain, whose morphology, size and location are estimated in order to determine possible malformations. Only 2 papers out of 11 clearly include twin pregnancies (mono- and dichorionic pregnancies) in their studies (Torrents-Barrena et al., 2020, 2021) for in-utero surgery planning; however, clinical and pathological cases involved are not balanced.

Few papers focus the attention on multi-organ analysis: the inter-organ relations are frequently exploited by doctors when navigating and interpreting medical images (Cerrolaza et al., 2019), making the multi-organ analysis essential for a good fetus evaluation. The lack of annotated datasets, even in this field, does not allow to (1) characterize the complex inter-organ relations and (2) underline the difference between pathological and physiological images since data representative of healthy and pathological cases is critical to develop CAD systems robust to pathologies and rare anatomies.

5. Biometry parameter estimation

Assessing fetal size and GA, as well as detecting fetal growth abnormalities, lay the foundation of modern prenatal care. Fetal biometry assessment represents the most common medical investigation undertaken in this regard. Before 14 weeks, the GA and the fetal size are estimated by the measurement of crown–rump length (CRL),
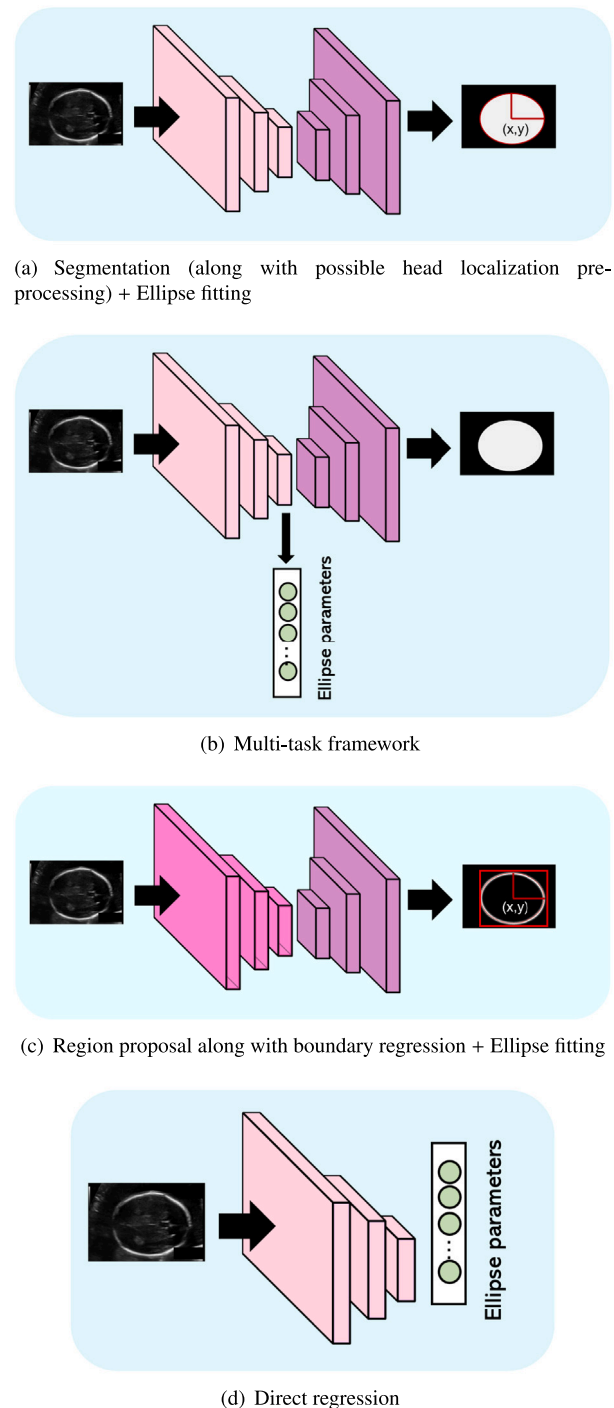
which is calculated from the top of fetus's head to the bottom of its torso. Once the CRL exceeds a fixed length (generally after 14 weeks), common measurements include HC, biparietal diameter (BPD), occipito-frontal diameter (OFD), trans-cerebellar diameter (TCD), lateral ventricles (LV), abdominal circumference (AC) and femur diaphysis length (FL). Fig. 8 shows image samples for each of these biometries. These fetal biometries are used to assess fetal growth trajectory and ensure normal fetal development when measured at different points in time (trimesters). Additionally, cardio-thoracic ratio (CTR) and cardiac axis biometrics are measured for diagnosing CHD.

Biometry measurements are performed in a standardized manner by identifying the appropriate sonographic plane and by precisely placing calipers in the corrected position (Salomon et al., 2019). Over the years, DL algorithms that provide automated placement of the calipers to measure BPD, OFD, CRL, TCD, HC and AC have been extensively exploited to reduce operator-dependent errors and improve accuracy of fetus well-being assessment.

Most of the approaches in literature tackle the problem of caliper measurement by segmenting the area of interest at first, as shown in Fig. 9a. In Sinclair et al. (2018), a VGG-like CNN is used to segment fetal head and successively provide HC and BPD measurements by means of ellipse fitting. More recently, encoder–decoder CNNs have been investigated for segmentation task. Here, the majority of published papers focus on HC measurement only, relying on the HC18 challenge dataset (Section 2.2).

The work in Rong et al. (2019) proposes an active-contour model guided by external forces that are derived with a U-Net-like CNN trained to segment the fetal head. For the same goal, in Skeika et al. (2020) a V-Net is used to automatically segment the fetal skull and the HC is extrapolated. V-Net is modified to encode attention mechanism in Zeng et al. (2021). A U-Net architecture is instead used in Aji et al. (2019). Similarly, in Qiao and Zulkernine (2020) a U-Net architecture is modified by adding dilated convolution layers after the last layer of the encoder and squeeze-and-excitation blocks on the skip connections. In Oghli et al. (2020) a multi-feature pyramid U-Net is proposed. An encoder–decoder network is also proposed in Bhalla et al. (2021). The network consists of dense blocks that strengthen feature propagation and channel attention in bottleneck to enhance the important features. In all these papers, HC is extrapolated by means of ellipse fitting methods.

Multi-task approaches following the scheme in Fig. 9b are also explored in the literature. The work in Sobhaninia et al. (2019) proposes a multi-task CNN inspired by the structure of LinkNet, for the automatic segmentation of fetal head and the successively estimation of the HC main axes, center and angle. The authors present another approach (Sobhaninia et al., 2020) based on a multi-scale and low complexity structure inspired by LinkNet. A multi-task approach is also exploited in Li et al. (2020), where head segmentation accuracy is enhanced by a feature pyramid inside a U-Net-like CNN. A regressor branch is added for predicting the HC value.

To improve segmentation accuracy, some papers incorporate a localization network in the pipeline to relieve the segmentation network from learning the position of the head. In Kim et al. (2019), a region proposal network is used as post processing procedure once the fetal head is segmented by U-Net on polar transformed images. HC and BPD are extrapolated by an ellipse fitting method. A Mask-RCNN is used in Al-Bander et al. (2019) to improve segmentation accuracy by incorporating an object localization framework. Coordinates are extrapolated by means of an ellipse fitting method.

A different strategy is used in Fiorentino et al. (2021), which considers the problem of fetal head caliper placement as a distance regression task, as shown in Fig. 9c. The proposed framework consists of a tiny-YOLO for head location and centering, and a regression CNN, which by means of distance field, accurately delineates the fetal head boundaries. HC is successively obtained using an ellipse fitting method. The work is further improved in Moccia et al. (2021) in which the framework is made end-to-end using a Mask-RCNN along with ellipse fitting for HC extrapolation. The direct regression of the contours coordinates is also performed in Meng et al. (2020d). However, CNN is jointly used with an attention refinement module along with a graph convolutional network (GCN). This approach does not requires any ellipse fitting method.

The work in Budd et al. (2019) further extends the research in the field of HC estimation by developing two probabilistic CNN methods: Monte Carlo Dropout during inference and a probabilistic U-Net. These methods are particularly useful in the clinical practice since multiple plausible semantic segmentation of fetal heads along with HC measurements are provided to the clinicians, which can choose the best option.

In order to avoid intermediate steps, such as segmentation CNN along with ellipse fitting methods, which may be computationally expensive (both for model training and labeling), the work in Zhang et al. (2020a) proposes a regression CNN to directly estimate the HC measure, as shown in Fig. 9d. Different CNNs are tested (including VGG16 and ResNet) as backbones. A similar approach is followed in Zhang et al. (2020b) in which saliency maps are used to provide interpretation of the regression CNN results.

Besides head biometries, a number of researchers is focusing on AC estimation. The work in Jang et al. (2017) proposes a method for AC estimation from 2D US data. The framework performs semantic segmentation of several anatomical structures (i.e., stomach bubble, amniotic fluid, and umbilical vein) by using a custom designed CNN. AC measurement is achieved through Hough transformation, and a plane acceptance check is performed by another CNN, which verifies the presence of all structures of interest. This method faces problems in predicting the AC precisely in the case of insufficient amniotic fluid, which commonly occurs when observing oversized fetuses. To improve the results achieved by Jang et al. (2017), the work in Kim et al. (2018) proposes a method consisting of a combination of multiple CNNs. A CNN is used to identify anatomical structures in US images (stomach bubble, amniotic fluid, and umbilical vein) and Hough transform is used to obtain an initial estimate of the AC. These data are fed to other CNNs to estimate the spine and bone positions, which are used to compute AC accurately. Then, a U-Net and a classification CNN are used to check whether the image is suitable for AC measurement.

Only one work (Cengiz and Yaqub, 2021) focuses on CRL measurement estimation. A U-Net is used to segment all fetus area and CRL is computed from the obtained segmentation.

Few papers in the literature focus on FL, LV biometry and CTR estimation. A first approach for automatically measuring the width of LV is proposed in Chen et al. (2020d). This approach relies on Mask-RCNN. FL measurement is performed with SegNet and image skeletonization in Zhu et al. (2021). The work in Chen et al. (2021a) focuses on CTR and cardiac axis estimation. The proposed CNN is a one-stage ellipse detection algorithm.

Multiple biometry estimation is addressed in Bano et al. (2021), where head, abdomen and femur are semantically segmented using state-of-the-art CNNs. This step is followed by region fitting and scale recovery for the biometry estimation. The work in Płotka et al. (2021) proposes a multi-task framework consisting of a U-Net along with ConvLSTM to jointly localize, classify and measure HC, AC and FL in fetal US video. A multiple biometry estimation is performed also in Ryou et al. (2019) with the aim of supporting first trimester fetal assessment from a single 3-D US scan. The framework automatically extracts a slice of the whole fetus in the sagittal view and simultaneously segments the whole fetus by means of a multi-task network. Automated segmentation of the whole fetus into head, abdomen and limbs is then performed to extrapolate biometry measurements. The work in Oghli et al. (2020) is further expanded in Oghli et al. (2021), in which HC, BPD, AC and FL measurements are also considered.

The work in Gao et al. (2021) proposes an approach that relies on unsupervised domain adaptation for HC and TCD biometry estimation from low cost US devices. The domain invariant representations for feature extraction are jointly learnt from both high-end and low cost US images and an unsupervised learning is used to calibrate the model in the low cost domain in order to produce consistent predictions. The final biometric estimation is performed fitting an ellipse to the contours (for HC measurements) and performing non-maximum suppression to find the pixels with greatest probabilities (for TCD measurements).

In Prieto et al. (2021), a fully-automated pipeline for GA estimation is proposed. The pipeline relies on BPD, HC, CRL, AC and FL estimation, which is accomplished following U-Net based structure segmentation.

An approach to automate the fetal head biometry estimation in real-time is proposed in Rasheed et al. (2021). Firstly, an AlexNet is used to classify and extract fetal head from US images; successively a U-Net is used to compute HC and BPD to estimate GA.

5.1. Limitations and open issues for biometry parameter estimation

Most of the papers dealing with biometry parameter estimation focus the attention on HC estimation only. Extensive research on HC parameter estimation is performed from 2D trans-ventricular planes, specifically concerning second and third trimesters screening. This may be attributed to the release of the HC18 dataset. From the virtuous example of the HC18 challenge, the research community should work to collect and release datasets for other biometries too, as very little has been done for other biometries such as CSP and CRL, instead. Table 6 summarizes the papers surveyed in this section.

However, the HC18 dataset does not allow studying advanced DL methodologies because of both limited number of images, anatomical structures (only fetal brain is shown) and consideration of healthy subjects only. Moreover, only one clinician participated in the annotation process, generating possible annotation bias. These issues must be addressed in order to develop a more robust research in the field. As regards parameter estimation methods, most of the papers in literature still rely on fitting method to extract object contours and successively measure biometry (as also shown in Fig. 9). The regression of object parameters would be a reasonable alternative since a single model to predict outputs directly from the inputs would improve the efficiency and the performance of the method. However, directly regressing single points from an input frame is highly non-linear and thus model convergence results to be more difficult to obtain Bulat and Tzimiropoulos (2016).

Considering possible future developments, an end-to-end approach to first identify the scan plane and then calculate biometries associated to that plane could be a valuable support tool for clinicians. Only the clinical study in Prieto et al. (2021) proposes something similar, but further research is needed.

Moreover, a unified framework to biometry estimation from multiple anatomical regions could also be explored. Here, the most peculiar challenge is the huge variability in terms of shape and morphology of the anatomical structures (i.e. abdomen less contrasted with the background as opposed to other organs, femur and cerebellum that do not fit an ellipse), as well as the dimension of the organs, which varies according to the gestational trimesters.

6. Other tasks

Besides standard-plane detection (Section 3), anatomical structure analysis (Section 4) and biometry measurements (Section 5), researchers are working on a number of emerging tasks, which are grouped and summarized in the following subsections.

6.1. Gender identification

Determining fetal gender by US scans is among the main tasks addressed during the early stages of pregnancy. With the aim of canceling this information from the US display to prevent unauthorized gender viewing, the work in Lakra et al. (2019) develops a residual network to identify frames containing gender-defining region in real-time.

6.2. Pose estimation

Studies about the structured description of the whole fetus in US are still rare. In Yang et al. (2019a), fetal pose is estimated by localizing 16 landmarks, including joints. The authors propose a self-supervised learning framework to fine tune a network to form visually plausible pose predictions: a pre-trained 3D U-Net predicts the heatmaps of the 16 landmarks with an intermediate fetal pose estimator. By retrieving a support set of atlases in the pose library via rigid registration, label proxies are produced to form the self-supervision. The landmark detector is tuned iteratively for on-line refinement and updated under the gradient checkpointing strategy if necessary.

6.3. Preterm-birth prediction

An early detection of preterm birth risks is crucial to timely intervene and preserve fetal survival. US data are typically inspected by expert doctors, which use hand-designed image features such as cervical length and anterior cervical angle. Lately, to overcome errors related to subjective method, DL was used to predict preterm birth.

In Włodarczyk et al. (2019), a U-Net is used to segment the cervix from US images. The masks obtained are used to estimate the cervical length and the anterior cervical angle. The cervical length is measured through the centerline algorithm and evaluated performing linear regression between estimated and ground truth cervix lengths. The anterior cervical angle, instead, is estimated using an iterative algorithm, that finds centroid point of the cervix mask and splits it into two shapes: the anterior cervical angle is measured between the anterior wall and the line between the last two centroids. A similar approach is proposed in Włodarczyk et al. (2020), which uses a U-Net with a parallel branch for classification to simultaneously segment and classify the cervix in 2D US images.

6.4. US simulation

US simulation through image synthesis has a wide range of applications, including medical procedure simulation, clinical training and multimodality image registration.

Using US imaging and optical tracking data acquired on a fetus phantom, Hu et al. (2017) simulate US images at given 3D spatial locations based on spatially-conditioned GANs, which sample anatomically accurate images conditionally on spatial position of the (real or mock) freehand US probe.

Besides the US-specific difficulties, synthesizing US images needs to be feasible for interactive use, which means to comply with limited computational resources. Therefore, Zhang et al. (2020c) proposes a patch-based GAN that aims to improve the quality of simulated US images while keeping computational time low, via image translation of computationally low-cost images to high quality simulation outputs. As additional input to the model, segmentation maps are provided to incorporate anatomical information potentially lost in low quality images, and also acoustic attenuation maps are given to better preserve acoustic shadows and directional artifacts information.

US image generation is explored also in Xu et al. (2020c), in which a cycle-consistent adversarial network (Cycle-GAN) is designed to simulate visually realistic third-trimester US images starting from unpaired second- and third-trimester FVSP US images.

6.5. Video summarization

Video summarization is crucial to lower the workload of clinicians when reviewing US examination. In Liu et al. (2020), DRL is applied to video summarization to select a subset of frames to create a shorter video that contains sufficient and essential information to facilitate retrospective analysis, using an encoder–decoder CNN structure to extract relevant features from frame sequences. Video summarization

is further investigated in Sharma et al. (2021), which proposes an approach for video description and clinical-workflow analysis during fetal US scans. The semi-automatic method to temporally segment the US videos into semantically meaningful segments is based on a double branch architecture in which a SonoNet is employed to extract spatial features, and an LSTM unit is used to model temporal dependency.

6.6. Shadow artifact removal

Acoustic shadows caused by sound-opaque occluders can potentially hide vital anatomical information in 2D US and thus can be a big burden for US analysis, ranging from anatomy segmentation to landmark detection. An automatic shadow detection method is presented in Meng et al. (2018). It generates a pixel-wise shadow confidence map from weakly labeled annotations, jointly using a FCN as shadow image discriminator, a feature attribution map from a GAN and an intensity saliency map from a graph cut model. The authors improve these results in Meng et al. (2019b), developing a CNN-based, weakly supervised method for automatic confidence estimation of shadow regions in 2D US images. By learning and transferring shadow features from weakly-labeled images, continuous shadow confidence maps are directly predicted from input images. Differently from previous work, in Meng et al. (2019a) a disentanglement method is presented to disjoint anatomical from shadow features, to generalize anatomical standard plane analysis for abnormality detection in early pregnancy. The authors proposed a multi-task architecture with adversarial training, evaluated on standard-plane/shadow artifacts classification tasks.

Together with the indiscriminate mixing of image properties, e.g. artifacts and anatomy, another challenge for DL algorithms in US image analysis is the presence of different acquisition devices characteristics. The work in Meng et al. (2020b), while considering the shadow artifacts presence, explores also the cross-device adaptation problem. The authors present a non-adversarial method that evaluates mutual information between latent features to disentangle categorical features and domain features in a semi-supervised learning framework.

6.7. Probe movement control

Accurate obstetric US scanning is highly operator dependent. This issue may be overcome with the use of US probe-movement guidance, especially to assist less-experienced operators. In Zhao et al. (2021), contrastive learning is used to train an end-to-end self-supervised network for visual-assisted probe movement via automated landmark retrieval. A set of landmarks is built on a virtual 3D fetal model, while during obstetric scanning a transformer is used to locate the nearest landmark through descriptor search between the current observation and the landmarks.

6.8. Volume reconstruction

In recent years, 3D US has become popular in prenatal care routine, commonly used to show a detailed picture of the baby and look at his face. However, 3D US could be useful for the assessment in standard planes of normal and abnormal structures, like fetal brain, spine, face, kidney and heart, to evaluate development and functionality of some organs, like heart and kidney, and to facilitate the identification of subtle fetal defects or malformations (Merz and Pashaj, 2017). However, 3D US probe is constrained by the limited field of view and poor operability (Mohamed and Siang, 2019). Thus, to address these problems, in Luo et al. (2021) an approach for fetal 3D freehand US reconstruction from 2D slices is proposed. The reconstruction process is approached with a convolutional LSTM. Self-supervised learning is applied on the reconstructed volumes to perform pseudo supervision and regularize the prediction of future frames. Finally, adversarial learning is introduced to improve the representation learning of anatomical shape priors, to prevent uneven reconstructions.
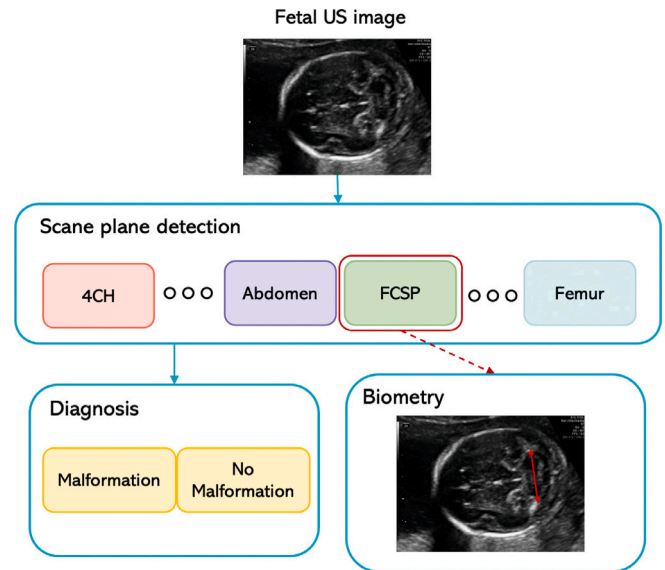

Fig. 10. Overall workflow of a computer-assisted tool to support fetal clinicians.

6.9. Limitations and open issues for other tasks

Section 6 presented a variety of emerging approaches in the field of DL for US image analysis. Some of these approaches address interesting tasks from the clinicians' perspective, including preterm-birth prediction and fetus pose estimation. Both may provide clinicians with decision support and, if further developed, CAD. Video summarization may also be seen as a task for directly supporting clinicians, as summarization is thought to relieve clinicians mental workload and provide automatic video reporting that can be included in the health records. Other approaches focus on more auxiliary tasks, including US simulation, shadow artifact removal and volume reconstruction. These approaches are needed to increase dataset size, improve image quality or allow fast 3D volume estimation from 2D US images, but do not present an explicit clinical application. From the DL algorithm point of view, it appears that the literature is still too sparse for a proper and fair comparison. We expect, however, that the research community will focus on these tasks in the next years.

7. Discussion and conclusions

This review analyzed a wide spectrum of the most recent DL algorithms for fetal US image analysis. Fetal US image analysis dates back to mid 1900s and a solid and rich literature today exists. Our survey started with the aim of answering the following questions:

Which are the most investigated tasks addressed using DL in the field of fetal US image analysis? According to our findings, standard plane detection (19.6%) and biometry parameter estimation (20.9%) are among the most investigated tasks (Fig. 2(a)). In standard plane detection, fetal brain, abdomen and heart standard planes are the most explored since biometric measurements and identification of abnormalities are mainly performed on these planes. The recognition of the specific anatomical landmarks is the key for evaluating the quality of each scanned plane and detection algorithms or attention mechanisms are thus particularly used in this domain. As regards biometry parameter estimation, the HC is the most investigated measurement. This is mainly due to the release of HC18 challenge, which allowed a more extensive research in the domain. Segmentation (along with possible head localization pre-processing) followed by fitting methods approaches are the most exploited ones. The 49.0% of the analyzed papers dealt with anatomical structure analysis. According to Fig. 7,

the 26,7% and the 20,0% of the papers work on fetal heart and brain applications, respectively. The cardiac examination of fetus in literature is designed to maximize the detection of heart anomalies and genetic syndromes. Detection rate is often optimized by recognizing the main fetal heart planes (4CH, LVOT and RVOT) at first. The analysis is frequently performed on 2D US images or integrating spatio-temporal information, which improves CNN performance since the temporal dependencies present in consecutive frames are exploited.

As regards fetal brain analysis, a great part of the papers surveyed focuses on the evaluation of brain structures. The analysis is performed both with 2D and, more recently, 3D images, but structure segmentation is mainly assessed with encoder–decoder inspired architectures. GA and brain development also occupy a consistent part of fetal brain evaluation. The anatomical regions that have received marginal attention are lungs, kidney and spine, whose approaches are also less varied.

Which are the main challenges in regard to fetal examination that are currently tackled by using DL? From our survey, it emerged that DL is today able to tackle US image-related challenges, including inhomogeneities, artifacts (i.e. shadows), poor contrast and intra- and inter-clinician acquisition and measurement variability, supporting clinicians in a wide range of clinical applications. For standard plane detection, current DL algorithms are robust when processing single-center datasets, collected and annotated by 1 or 2 experts mostly from a single US machine. The algorithms are able to tackle challenges such as a high variability (size and shape) of fetuses among different GAs, as well as the presence of shared anatomical structures among different standard planes. Here, a winning strategy in the literature is to check that a predefined number of structures appears in the US probe field of view.

As regards anatomical-organ contest, DL has proven to be an essential tool to support the segmentation process of different anatomical structures, which may vary in shape and size over GAs. Boundary incompleteness is a common problem for those anatomical structures that are not particularly contrasted with respect to the background. Up to now, all the DL segmentation techniques, which are mainly supervised, outperformed other state-of-the-art methods in both performance and speed. However, these approaches may be unable to deal with boundary information loss due to a lack of prior shape information. Dilated convolutions along with residual learning are, thus, often exploited to help to increase CNN field of view and to recover information loss, respectively. CAD systems that make full use of DL have been explored, for the past few decades, as another research topic in the field, helping clinicians in their decision-making process. For biometry-parameter estimation, DL today achieves robust results on HC estimation tackling challenges such as different positions of the head in the image, varying dimensions of fetal head among the gestational trimesters and partially visible head skull. These challenges are today tackled for still images acquired from a single US machine and annotated by a limited number of clinicians. Only one paper proposed an approach in the field invariant to significant image distribution shift between image types (Gao et al., 2021). Besides tasks addressed in Sections 3, 4, and 5, recently other tasks have been analyzed in Section 6. Emerging fields such as preterm birth prediction or fetal pose estimation are investigated to intervene and preserve fetal survival and quantify fetal movements. Techniques which improve image quality and image analysis are also exploited, such as shadow artifact removal and 3D volume reconstruction from 2D slices, which is particularly useful to provide information on the entire anatomy (typical of 3D imaging) together with a more complete field of view.

Are the commonly-used datasets sufficient enough for robust DL algorithm development and testing? Collecting and sharing annotated datasets for DL algorithm development is a well-known problem for the medical-image analysis community. Up to now, only few datasets are available in the US fetal field (Section 2.2). Besides practical issues (annotating data is a labor-intensive process), there are social and ethical concerns that have to be considered. To attenuate these

issues, all these aspects should be deeply studied in collaboration with non-technical communities (e.g., lawyers, ethicists, ...). This step is crucial with a view to develop robust and generalizable algorithms.

In terms of number of images, almost all articles were evaluated with less than 1000 images and some of them even with less then 300 images. Moreover, the number of images that have been used are missing in some of the articles. In order to have a high degree of robustness, train/test size must be always clarified as well as patient number, which is an essential information to avoid that the same woman' scan is part of both training/validation and test sets. A cross-validation is strongly suggested when limited data sample is available, especially if we are facing with multi-class problems. With cross-validation strategies, all the data is used to evaluate model capability, making algorithm performance more reliable. Providing information about GA distribution is another requirement: fetus changes and grows during gestation time, thus also the evaluation of its status varies among trimesters. Routine US examinations are mainly performed during second trimester of gestation (Salomon et al., 2022); this reflects also on the analyzed papers (Tables 4, 5, 6). The majority of them includes data acquired during 18 to 22 weeks, which serves as a baseline to evaluate fetal growth and healthy status and to perform a timely detection of major congenital anomalies. Hence, the analysis of most anatomical structures is performed on US scans of this GA and of third trimester, to evaluate and monitor compliance with standard fetal organ sizes. The only exception is the estimation of position and echostructure of the placenta, which is mainly performed during first trimester (Hu et al., 2019, 2021; Oguz et al., 2018; Looney et al., 2018, 2017; Yang et al., 2017; Schwartz et al., 2021). As regards biometry parameter estimation, the vast majority of papers deal with measurements performed during mid and third trimesters (e.g. HC, AC, FL) as opposed to first trimester ones (e.g. CRL), Cengiz and Yaqub (2021) and Ryou et al. (2019), which is however of fundamental importance to estimate potential due date. Once the fetus has passed 13 + 6 weeks, all the other measurements are performed to determine how the baby is progressing.

Which are open issues that still have to be addressed by DL in the field? Hereafter, we discuss about the main open issues and future research directions that we have identified in the field of fetal US image analysis.

Multi-expert image annotation. The lack of image annotation by multiple clinicians currently represents a major drawback. While it is undeniable that having more than one expert annotator is expensive in terms of money and time, this multi-expert annotation is crucial for the development of robust DL algorithms and for fair comparison of their performances. Currently only few papers provide a measure of the degree of intra- and inter-clinician variability. However, this parameter should be assessed as it is an indicator of the degree of image complexity. Hence, it would be advisable to always consider multiple annotations from several observers and from the same observer at different times. This is particularly crucial when performing (1) biometry parameters estimation, since measurements are performed at pixel level requiring operator independence to be reliable; (2) CAD in anatomical structure analysis, which should not depend on only one observer/observation. With these considerations in mind, results interpretation must be carefully evaluated when manual annotation is used as the main reference for the study.

Performance evaluation. The absence of a systematic evaluation workflow emerged as a critical point. Fair algorithm comparison is often hampered due to an inconsistent use of performance metrics and/or testing datasets. A direct comparison of methods, indeed, could not be performed for all the sections under exam, since each work is validated through different measures. Moreover, despite the efforts of some researchers and International Organizations, the publicly available datasets are still too few and often released for some specific tasks (e.g., biometric measurements in a body district). The lack of available datasets in fetal anatomical structure analysis is, in fact, still to fill:

addressing this issue means to increase the research in the field to a certain level, thus, providing a better evaluation of different aspects (from GA estimation up to fetal well-being assessment). DL algorithms show good performance in controlled evaluation setup, where often data from one single clinical center are available. This does not reflect the actual practice, where different US devices and probes are used, and may result in poor generalization. As for now, only few papers deal with domain generalization, proving that further research is needed.

2D data stream analysis. The majority of the selected papers neglects the temporal dimension and treats each image individually. This may be considered as an issue for standard plane detection, as clinicians identify the standard plane by manoeuvring the US probe to the desired scan plane, which results in a 2D data stream. Furthermore, processing still images may result in poor decision support to clinicians in cases in which it is important to evaluate dynamic processes, such as heart contractility analysis and fetus movement assessment.

Comprehensive analysis. Developing comprehensive computational models of the fetus is particularly challenging: as the baby grows, the fetus' body is in constant transformation, with significant structural and physiological changes among trimesters leading to an important inter- and intra-organ variability. The creation of anatomically accurate models able to characterize the complexity of fetus anatomy represents, thus, one of the biggest challenges. In addition, the acquisition of quality images is the first step to adequately assess fetus well-being: an end-to-end approach able to correctly classify fetal planes and successively evaluate biometries or fetus defect has not been fully exploited yet (see Fig. 10). Among the few papers that tried to create an unified approach, the work in Arnaout et al. (2021) is the most complete. However, the application is limited to only one anatomical structure (in this case the heart), and thus it does not exploit the potential of DL for distinguishing in size and shape different objects.

Computer-assisted diagnosis. In CAD, DL algorithms are used to analyze images from a patient population and build a model capable of extracting information related to a certain disease and eventually provide a disease outcome prediction. Therefore, CAD may be used as a support tool in the clinical decision making process. However, data coming from pathological subjects are usually more difficult to collect compared to the ones coming from healthy subjects and building robust models performing CAD results particularly challenging. This emerged also from surveyed papers, among which only few of them included pathological subjects data, mainly related to fetal congenital heart disease and fetal brain malformations. Hence, in function of a diagnosis it is necessary to analyze fetal anatomical structures including also a wide range of pathological cases while considering also the corresponding GA information. In this perspective, researchers should work on creating balanced datasets, possibly exploring generative approaches or multicentric studies.

Semi, weak and self-supervised learning. To attenuate the issue of having small annotated datasets, researchers in closer fields are proposing semi, weakly or self-supervised approaches. In the fetal US domain, only 13 out 153 papers investigate the potential of such approaches. These include (Gao et al., 2020; Tan et al., 2019) to classify and detect planes, and Meng et al. (2020b) to explore cross-device adaptation problems. Weak supervision tasks are involved in scan plane detection (Baumgartner et al., 2017), fetal heart representations (Gao and Noble, 2017), multi-organ analysis (Gao and Noble, 2019) and shadow detection (Meng et al., 2018, 2019b). Self supervised techniques are used for scan plane detection (Chen et al., 2019a), fetal pose estimation (Yang et al., 2019a), probe movement estimation (Zhao et al., 2021), fetal 3D reconstruction (Luo et al., 2021). Other interesting approaches propose unsupervised learning techniques, such as in Chen et al. (2020c) for multi-organ classification and Gao et al. (2021) for biometry estimation in low cost US devices. Even if there is a recent attempt to exploit such techniques in the field, still too few papers make use of them, proving that this is still an unexplored field in fetal US domain.

Multimodal image analysis. Integrating US with other imaging modalities may be beneficial to provide clinicians with decision support and context awareness. However, few papers in the literature addressed this topic. Multimodal image analysis may be exploited for surgical planning and guidance, for example enriching the information provided to clinicians using mixed reality, as suggested in Torrents-Barrena et al. (2021).

Model efficiency. The computational cost associated to training and deployment of DL models is not reported in the majority of the surveyed papers. An high computational cost hampers the efficient deployment of the DL algorithms on single-board computers for on-the-edge computation. At the same time, a high computational cost is associated with a high CO_2 consumption. Model efficiency is a critical aspect that is currently monitored by International Organizations[6] and performance on this regard should be considered as additional evaluation metric.

On-device DL in fetal US. All the world's tech giants embrace DL to provide next-level products and services in fields such social networks, self-driving cars and finance. DL applications in medical imaging give great results in research papers, however their commercial use is not common yet. Among the DL algorithms used for commercial purposes in this field, SonoLyst is the first fully integrate AI tool in the world that allows the identification of 20 views recommended by the ISUOG mid-trimester practice guidelines for fetal sonography imaging.[7]

Use of federated learning. The majority of the surveyed papers relies on single-center datasets or datasets made available through international initiatives (e.g., Grand Challenge, Section 2.2). Accessing sufficiently large and diverse datasets of fetal US images is still a significant challenge. The collaboration among clinical centers based on centrally-shared US images face privacy and ownership concerns. Federated learning is a novel paradigm for data-private multi-institutional collaborations, where model-learning leverages all available data without sharing data between institutions (Sheller et al., 2020). Despite the benefits of such a paradigm, none of the reviewed articles has implemented it. This key aspect represents a starting point for future development.

Adherence to the ethics guidelines for trustworthy AI. With the rapid development of DL algorithms for fetal US image analysis, the application of ethical principles and guidelines has become crucial. In 2018, the European Commission has published a white paper[8] titled "Ethics Guidelines for Trustworthy AI" to stress on the important of respecting and promoting ethical principles in all the steps that involve the use of DL, from design to deployment. However, no attempts have been made in the reviewed papers in this directions. The Guidelines present an assessment list that offers guidance on each requirement's practical implementation. This assessment list should be filled by researchers and reported as supplementary material.

Advances in DL for fetal US analysis have demonstrated potential to support clinicians in clinical routine. From one side, DL algorithms may automatize repetitive tasks such as automatic scan plane detection and biometry estimation, allowing clinicians to save time and to better concentrate on image interpretation (Matthew et al., 2022); intra-clinician variability may also be reduced, enabling a more objective examination. From the other side, there is a growing interest in DL algorithms for complementing the current clinical protocols and enabling a more thorough examination in the future. Examples in this direction include fetal anomaly detection and 3D placenta analysis with mixed reality.

To conclude, this review introduced and discussed the most innovative and effective DL methods found in the literature for fetal US image

---

[6] https://cordis.europa.eu/programme/id/H2020-EU.3.5.

[7] https://www.intelligentultrasound.com/scannav-assist/.

[8] https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html.

analysis. The methods were summarized using tables reporting performance metrics, training set and test set sizes, number of annotators and related GAs. For each section, the most relevant pros and cons of each group of approaches were pointed out. We hope our review may be helpful for young researchers to get a better picture of the methods available, leading to a speed-up on the development and enhancement of methods for fetal US image analysis.

Nomenclature

AC:	Abdominal Circumference
Acc:	Accuracy
AUC:	Area Under the ROC Curve
BPD:	Biparietal Diameter
CAD:	Computer-Assisted Diagnosis
CNN:	Convolutional Neural Networks
CRL:	Crown–Rump Length
CSP:	Cavum Septum Pellucidum
CTR:	Cardio-Thoracic Ratio
DL:	Deep Learning
DRL:	Deep Reinforcement Learning
DSC:	Dice Similarity Coefficient
ED:	Euclidean Distance
FASP:	Fetal Abdomen Standard Plane
FBSP:	Fetal Brain Standard Plane
FCN:	Fully Convolutional Networks
FCSP:	Fetal Trans-Cerebellum Standard Plane
FFASP:	Fetal Face Axial Standard Plane
FFSP:	Fetal Facial Standard Plane
FFESP:	Fetal Femur Standard Plane
FL:	Femur Diaphysis Length
FLVSP:	Fetal Lumbosacral Spine Standard Plane
FTSP:	Fetal Trans-Thalamic Standard Plane
FVSP:	Fetal Trans-Ventricular Standard Plane
FV:	Fetal Ventriculomegaly
*F*1:	F1-score
GA:	Gestational Age
GAN:	Generative Adversarial Network
GRU-RCN:	Gated-Recurrent-Unit Recurrent Convolutional Network
HC:	Head Circumference
HD:	Hausdorff Distance
IoU:	Intersection over Union
ISBI:	International Symposium on Biomedical Images
ISUOG:	International Society of Ultrasound in Obstetrics and Gynecology
KLD:	Kullback–Leibler Divergence
LVOT:	Left Ventricular Outflow Tract
LVR:	Lateral Ventricle Ratio
LSTM:	Long Short-Term Memory
MAE:	Mean Absolute Error
MICCAI:	Medical Image Computing and Computer Assessed Intervention
MVP:	Maximum Vertical Pocket
NAS:	Neural Architecture Search
NSS:	Normalized Scanpath Saliency
OFD:	Occipito-Frontal Diameter
Prec:	Precision
Rec:	Recall
RMSE:	Root Mean Squared Error
RNN:	Recurrent Neural Network
ROC:	Receiver Operating Characteristic
RVOT:	Right Ventricular Outflow Tract
Spec:	Specificity
SSD:	Single-Shot Detector
SSL:	Self-Supervised Learning
SVM:	Support-Vector Machine
TCD:	Trans-Cerebellar Diameter
US:	Ultrasound
3VT:	Three-Vessel Trachea
3VV:	Three-Vessel View
4CH:	Four Chamber View

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Aji, C.P., Fatoni, M.H., Sardjono, T.A., 2019. Automatic measurement of fetal head circumference from 2-dimensional ultrasound. In: 2019 International Conference on Computer Engineering, Network, and Intelligent Multimedia. IEEE, pp. 1–5.

Akkus, Z., Cai, J., Boonrod, A., Zeinoddini, A., Weston, A.D., Philbrick, K.A., Erickson, B.J., 2019. A survey of deep-learning applications in ultrasound: Artificial intelligence–powered ultrasound for improving clinical workflow. J. Am. Coll. Radiol. 16 (9), 1318–1328.

Al-Bander, B., Alzahrani, T., Alzahrani, S., Williams, B.M., Zheng, Y., 2019. Improving fetal head contour detection by object localisation with deep learning. In: Annual Conference on Medical Image Understanding and Analysis. Springer, pp. 142–150.

Alsharid, M., El-Bouri, R., Sharma, H., Drukker, L., Papageorghiou, A.T., Noble, J.A., 2020. A curriculum learning based approach to captioning ultrasound images. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. Springer, pp. 75–84.

An, S., Zhu, H., Wang, Y., Zhou, F., Zhou, X., Yang, X., Zhang, Y., Liu, X., Jiao, Z., He, Y., 2021. A category attention instance segmentation network for four cardiac chambers segmentation in fetal echocardiography. Comput. Med. Imaging Graph. 93, 101983.

Arnaout, R., Curran, L., Zhao, Y., Levine, J.C., Chinn, E., Moon-Grady, A.J., 2021. An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. Nat. Med. 27 (5), 882–891.

Aydin, O.U., Taha, A.A., Hilbert, A., Khalil, A.A., Galinovic, I., Fiebach, J.B., Frey, D., Madai, V.I., 2021. On the usage of average hausdorff distance for segmentation performance assessment: hidden error when used for ranking. Eur. Radiol. Exp. 5 (1), 1–7.

Bano, S., Dromey, B., Vasconcelos, F., Napolitano, R., David, A.L., Peebles, D.M., Stoyanov, D., 2021. AutoFB: Automating fetal biometry estimation from standard ultrasound planes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 228–238.

Baumgartner, C.F., Kamnitsas, K., Matthew, J., Fletcher, T.P., Smith, S., Koch, L.M., Kainz, B., Rueckert, D., 2017. SonoNet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. IEEE Trans. Med. Imaging 36 (11), 2204–2215.

Bhalla, P., Sunkaria, R.K., Kamboj, A., Bedi, A.K., 2021. Automatic fetus head segmentation in ultrasound images by attention based encoder decoder network. In: 2021 12th International Conference on Computing Communication and Networking Technologies. IEEE, pp. 1–7.

Bijma, H.H., van der Heide, A., Wildschut, H.I., 2008. Decision-making after ultrasound diagnosis of fetal abnormality. Reprod. Health Matters 16 (31), 82–89.

Bohlender, S., Oksuz, I., Mukhopadhyay, A., 2021. A survey on shape-constraint deep learning for medical image segmentation. IEEE Rev. Biomed. Eng..

Budd, S., Sinclair, M., Khanal, B., Matthew, J., Lloyd, D., Gomez, A., Toussaint, N., Robinson, E.C., Kainz, B., 2019. Confident head circumference measurement from ultrasound with real-time feedback for sonographers. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 683–691.

Bulat, A., Tzimiropoulos, G., 2016. Human pose estimation via convolutional part heatmap regression. In: European Conference on Computer Vision. Springer, pp. 717–732.

Burgos-Artizzu, X.P., Coronado-Gutiérrez, D., Valenzuela-Alcaraz, B., Bonet-Carne, E., Eixarch, E., Crispi, F., Gratacós, E., 2020. Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. Sci. Rep. 10 (1), 1–12.

Bushra, S.N., Shobana, G., 2021. Obstetrics and gynaecology ultrasound image analysis towards cryptic pregnancy using deep learning-a review. In: 5th International Conference on Intelligent Computing and Control Systems. IEEE, pp. 949–953.

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., Durand, F., 2018. What do different evaluation metrics tell us about saliency models? IEEE Trans. Pattern Anal. Mach. Intell. 41 (3), 740–757.

Cai, Y., Droste, R., Sharma, H., Chatelain, P., Drukker, L., Papageorghiou, A.T., Noble, J.A., 2020. Spatio-temporal visual attention modelling of standard biometry plane-finding navigation. Med. Image Anal. 65, 101762.

Cai, Y., Sharma, H., Chatelain, P., Noble, J.A., 2018. Multi-task sonoeyenet: detection of fetal standardized planes assisted by generated sonographer attention maps. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 871–879.

Casella, A., Moccia, S., Paladini, D., Frontoni, E., De Momi, E., Mattos, L.S., 2021. A shape-constraint adversarial framework with instance-normalized spatio-temporal features for inter-fetal membrane segmentation. Med. Image Anal. 70, 102008.

Cengiz, S., Yaqub, M., 2021. Automatic fetal gestational age estimation from first trimester scans. In: International Workshop on Advances in Simplifying Medical Ultrasound. Springer, pp. 220–227.

Cerrolaza, J.J., Picazo, M.L., Humbert, L., Sato, Y., Rueckert, D., Ballester, M.Á.G., Linguraru, M.G., 2019. Computational anatomy for multi-organ analysis in medical imaging: A review. Med. Image Anal. 56, 44–67.

Cerrolaza, J.J., Sinclair, M., Li, Y., Gomez, A., Ferrante, E., Matthew, J., Gupta, C., Knight, C.L., Rueckert, D., 2018. Deep learning with ultrasound physics for fetal skull segmentation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging. pp. 564–567.

Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., Rueckert, D., 2019a. Self-supervised learning for medical image analysis using image context restoration. Med. Image Anal. 58, 101539.

Chen, P., Chen, Y., Deng, Y., Wang, Y., He, P., Lv, X., Yu, J., 2020b. A preliminary study to quantitatively evaluate the development of maturation degree for fetal lung based on transfer learning deep model from ultrasound images. Int. J. Comput. Assist. Radiol. Surg. 15 (8), 1407–1415.

Chen, Y., Fan, H., Xu, B., Yan, Z., Kalantidis, Y., Rohrbach, M., Yan, S., Feng, J., 2019b. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In: IEEE/CVF International Conference on Computer Vision. pp. 3435–3444.

Chen, X., He, M., Dan, T., Wang, N., Lin, M., Zhang, L., Xian, J., Cai, H., Xie, H., 2020d. Automatic measurements of fetal lateral ventricles in 2D ultrasound images using deep learning. Front. Neurol. 11, 526.

Chen, Z., Liu, Z., Du, M., Wang, Z., 2021c. Artificial intelligence in obstetric ultrasound: An update and future applications. Front. Med. 1431.

Chen, Q., Liu, Y., Hu, Y., Self, A., Papageorghiou, A., Noble, J.A., 2020c. Cross-device cross-anatomy adaptation network for ultrasound video analysis. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. Springer, pp. 42–51.

Chen, L., Tian, Y., Deng, Y., 2021b. Neural network algorithm-based three-dimensional ultrasound evaluation in the diagnosis of fetal spina bifida. Sci. Program. 2021.

Chen, H., Wu, L., Dou, Q., Qin, J., Li, S., Cheng, J.-Z., Ni, D., Heng, P.-A., 2017. Ultrasound standard plane detection using a composite neural network framework. IEEE Trans. Cybern. 47 (6), 1576–1586.

Chen, C., Yang, X., Huang, R., Shi, W., Liu, S., Lin, M., Huang, Y., Yang, Y., Zhang, Y., Luo, H., et al., 2020a. Region proposal network with graph prior and IoU-balance loss for landmark detection in 3D ultrasound. In: 2020 IEEE 17th International Symposium on Biomedical Imaging. IEEE, pp. 1–5.

Chen, J., Zhang, Y., Wang, J., Zhou, X., He, Y., Zhang, T., 2021a. EllipseNet: Anchor-free ellipse detection for automatic cardiac biometrics in fetal echocardiography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 218–227.

Cho, H.C., Sun, S., Hyun, C.M., Kwon, J.-Y., Kim, B., Park, Y., Seo, J.K., 2021. Automated ultrasound assessment of amniotic fluid index using deep learning. Med. Image Anal. 69, 101951.

Dias, T., Sairam, S., Kumarasiri, S., 2014. Ultrasound diagnosis of fetal renal abnormalities. Best Pract. Res. Clin. Obstet. Gynaecol. 28 (3), 403–415.

Diniz, P.H., Yin, Y., Collins, S., 2020. Deep learning strategies for ultrasound in pregnancy. Eur. Med. J. 6 (1), 73.

Dong, J., Liu, S., Liao, Y., Wen, H., Lei, B., Li, S., Wang, T., 2019a. A generic quality control framework for fetal ultrasound cardiac four-chamber planes. IEEE J. Biomed. Health Inf. 24 (4), 931–942.

Dong, J., Liu, S., Wang, T., 2019b. ARVBNet: Real-time detection of anatomical structures in fetal ultrasound cardiac four-chamber planes. In: Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting. Springer, pp. 130–137.

Dou, H., Yang, X., Qian, J., Xue, W., Qin, H., Wang, X., Yu, L., Wang, S., Xiong, Y., Heng, P.-A., et al., 2019. Agent with warm start and active termination for plane localization in 3D ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 290–298.

Dozen, A., Komatsu, M., Sakai, A., Komatsu, R., Shozu, K., Machino, H., Yasutomi, S., Arakaki, T., Asada, K., Kaneko, S., et al., 2020. Image segmentation of the ventricular septum in fetal cardiac ultrasound videos based on deep learning using time-series information. Biomolecules 10 (11), 1526.

Droste, R., Cai, Y., Sharma, H., Chatelain, P., Papageorghiou, A.T., Noble, J.A., 2019. Towards capturing sonographic experience: cognition-inspired ultrasound video saliency prediction. In: Annual Conference on Medical Image Understanding and Analysis. Springer, pp. 174–186.

El Jurdi, R., Petitjean, C., Honeine, P., Cheplygina, V., Abdallah, F., 2021. High-level prior-based loss functions for medical image segmentation: A survey. Comput. Vis. Image Underst. 210, 103248.

Fiorentino, M.C., Moccia, S., Capparuccini, M., Giamberini, S., Frontoni, E., 2021. A regression framework to head-circumference delineation from US fetal images. Comput. Methods Programs Biomed. 198, 105771.

Franz, A., Schmidt-Richberg, A., Orasanu, E., Lorenz, C., 2021. Deep learning-based spine centerline extraction in fetal ultrasound. In: Bildverarbeitung Für Die Medizin 2021. Springer, pp. 263–268.

Gao, Y., Beriwal, S., Craik, R., Papageorghiou, A.T., Noble, J.A., 2020. Label efficient localization of fetal brain biometry planes in ultrasound through metric learning. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. Springer, pp. 126–135.

Gao, Y., Lee, L., Droste, R., Craik, R., Beriwal, S., Papageorghiou, A., Noble, A., 2021. A dual adversarial calibration framework for automatic fetal brain biometry. In: IEEE/CVF International Conference on Computer Vision. pp. 3246–3254.

Gao, Y., Noble, J.A., 2017. Detection and characterization of the fetal heartbeat in free-hand ultrasound sweeps with weakly-supervised two-streams convolutional networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 305–313.

Gao, Y., Noble, J.A., 2019. Learning and understanding deep spatio-temporal representations from free-hand fetal ultrasound sweeps. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 299–308.

Garcia-Canadilla, P., Sanchez-Martinez, S., Crispi, F., Bijnens, B., 2020. Machine learning in fetal cardiology: what to expect. Fetal Diagn. Ther. 47 (5), 363–372.

Gong, Y., Zhang, Y., Zhu, H., Lv, J., Cheng, Q., Zhang, H., He, Y., Wang, S., 2019. Fetal congenital heart disease echocardiogram screening based on DGACNN: adversarial one-class classification combined with video transfer learning. IEEE Trans. Med. Imaging 39 (4), 1206–1222.

He, S., Lin, Z., Yang, X., Chen, C., Wang, J., Shuang, X., Deng, Z., Liu, Q., Cao, Y., Lu, X., et al., 2021. Statistical dependency guided contrastive learning for multiple labeling in prenatal ultrasound. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 190–198.

Hermawati, F.A., Tjandrasa, H., Suciati, N., 2021. Phase-based thresholding schemes for segmentation of fetal thigh cross-sectional region in ultrasound images. J. King Saud Univ.-Comput. Inf. Sci..

Hesse, L.S., Namburete, A.I.L., 2020. Improving u-net segmentation with active contour based label correction. In: Annual Conference on Medical Image Understanding and Analysis.

van den Heuvel, T.L., de Bruijn, D., de Korte, C.L., Ginneken, B.v., 2018. Automated measurement of fetal head circumference using 2D ultrasound images. PLoS One 13 (8), e0200412.

Hu, Y., Gibson, E., Lee, L.-L., Xie, W., Barratt, D.C., Vercauteren, T., Noble, J.A., 2017. Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks. In: Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment. Springer, pp. 105–115.

Hu, Z., Hu, R., Yan, R., Mayer, C., Rohling, R.N., Singla, R., 2021. Automatic placenta abnormality detection using convolutional neural networks on ultrasound texture. In: Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis. Springer, pp. 147–156.

Hu, R., Singla, R., Yan, R., Mayer, C., Rohling, R.N., 2019. Automated placenta segmentation with a convolutional neural network weighted by acoustic shadow detection. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 6718–6723.

Huang, W., Bridge, C.P., Noble, J.A., Zisserman, A., 2017. Temporal HeartNet: towards human-level automatic analysis of fetal cardiac screening video. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 341–349.

Huang, R., Xie, W., Noble, J.A., 2018. VP-Nets: Efficient automatic localization of key brain structures in 3D fetal neurosonography. Med. Image Anal. 47, 127–139.

Huang, Q., Zeng, Z., 2017. A review on real-time 3D ultrasound imaging technology. BioMed Res. Int. 2017.

Ishikawa, G., Xu, R., Ohya, J., Iwata, H., 2019. Detecting a fetus in ultrasound images using grad CAM and locating the fetus in the uterus.. In: International Conference on Pattern Recognition Applications and Methods. pp. 181–189.

Jang, J., Park, Y., Kim, B., Lee, S.M., Kwon, J.-Y., Seo, J.K., 2017. Automatic estimation of fetal abdominal circumference from ultrasound images. IEEE J. Biomed. Health Inf. 22 (5), 1512–1520.

Judd, T., Ehinger, K., Durand, F., Torralba, A., 2009. Learning to predict where humans look. In: 2009 IEEE 12th International Conference on Computer Vision. IEEE, pp. 2106–2113.

Kim, B., Kim, K.C., Park, Y., Kwon, J.-Y., Jang, J., Seo, J.K., 2018. Machine-learning-based automatic identification of fetal abdominal circumference from ultrasound images. Physiol. Meas. 39 (10), 105007.

Kim, H.P., Lee, S.M., Kwon, J.-Y., Park, Y., Kim, K.C., Seo, J.K., 2019. Automatic evaluation of fetal head biometry from ultrasound images using machine learning. Physiol. Meas. 40 (6), 065009.

Komatsu, M., Sakai, A., Komatsu, R., Matsuoka, R., Yasutomi, S., Shozu, K., Dozen, A., Machino, H., Hidaka, H., Arakaki, T., et al., 2021. Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning. Appl. Sci. 11 (1), 371.

Kong, P., Ni, D., Chen, S., Li, S., Wang, T., Lei, B., 2018. Automatic and efficient standard plane recognition in fetal ultrasound images via multi-scale dense networks. In: Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis. Springer, pp. 160–168.

Lakra, P.P., Kumar, A., Mohanram, N., Krishnamurthi, G., Thittai, A.K., 2019. Deep-learning based identification of frames containing Foetal Gender Region during early second trimester ultrasound scanning. In: 2019 IEEE International Ultrasonics Symposium. IEEE, pp. 471–474.

Lee, L.H., Bradburn, E., Papageorghiou, A.T., Noble, J.A., 2020. Calibrated Bayesian neural networks to estimate gestational age and its uncertainty on fetal brain ultrasound images. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. Springer, pp. 13–22.

Lee, L.H., Gao, Y., Noble, J.A., 2021. Principled ultrasound data augmentation for classification of standard planes. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 729–741.

Li, Y., Khanal, B., Hou, B., Alansary, A., Cerrolaza, J.J., Sinclair, M., Matthew, J., Gupta, C., Knight, C., Kainz, B., et al., 2018. Standard plane detection in 3D fetal ultrasound using an iterative transformation network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 392–400.

Li, Y., Xu, R., Ohya, J., Iwata, H., 2017. Automatic fetal body and amniotic fluid segmentation from fetal ultrasound images by encoder-decoder network with inner layers. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 1485–1488.

Li, P., Zhao, H., Liu, P., Cao, F., 2020. Automated measurement network for accurate segmentation and parameter modification in fetal head ultrasound images. Med. Biol. Eng. Comput. 58 (11), 2879–2892.

Liang, J., Huang, R., Kong, P., Li, S., Wang, T., Lei, B., 2019. SPRNet: Automatic fetal standard plane recognition network for ultrasound images. In: Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis. Springer, pp. 38–46.

Lin, Z., Le, M.H., Ni, D., Chen, S., Li, S., Wang, T., Lei, B., 2018. Quality assessment of fetal head ultrasound images based on faster r-CNN. In: Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation. Springer, pp. 38–46.

Lin, Z., Li, S., Ni, D., Liao, Y., Wen, H., Du, J., Chen, S., Wang, T., Lei, B., 2019. Multi-task learning for quality assessment of fetal head ultrasound images. Med. Image Anal. 58, 101548.

Liu, Z., Huang, X., Yang, X., Gao, R., Li, R., Zhang, Y., Huang, Y., Zhou, G., Xiong, Y., Frangi, A.F., et al., 2021. Generalize ultrasound image segmentation via instant and plug & play style transfer. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 419–423.

Liu, T., Meng, Q., Vlontzos, A., Tan, J., Rueckert, D., Kainz, B., 2020. Ultrasound video summarization using deep reinforcement learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 483–492.

Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S.X., Ni, D., Wang, T., 2019. Deep learning in medical ultrasound analysis: a review. Engineering 5 (2), 261–275.

Looney, P., Stevenson, G.N., Nicolaides, K.H., Plasencia, W., Molloholli, M., Natsis, S., Collins, S.L., 2017. Automatic 3D ultrasound segmentation of the first trimester placenta using deep learning. In: 2017 IEEE 14th International Symposium on Biomedical Imaging. IEEE, pp. 279–282.

Looney, P., Stevenson, G.N., Nicolaides, K.H., Plasencia, W., Molloholli, M., Natsis, S., Collins, S.L., 2018. Fully automated, real-time 3D ultrasound segmentation to estimate first trimester placental volume using deep learning. JCI Insight 3 (11).

Luo, M., Yang, X., Huang, X., Huang, Y., Zou, Y., Hu, X., Ravikumar, N., Frangi, A.F., Ni, D., 2021. Self context and shape prior for sensorless freehand 3D ultrasound reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 201–210.

Luyckx, V.A., Brenner, B.M., 2015. Birth weight, malnutrition and kidney-associated outcomes—a global concern. Nat. Rev. Nephrol. 11 (3), 135–149.

Matthew, J., Skelton, E., Day, T.G., Zimmer, V.A., Gomez, A., Wheeler, G., Toussaint, N., Liu, T., Budd, S., Lloyd, K., et al., 2022. Exploring a new paradigm for the fetal anomaly ultrasound scan: Artificial intelligence in real time. Prenat. Diagn. 42 (1), 49–59.

Meng, Q., Baumgartner, C., Sinclair, M., Housden, J., Rajchl, M., Gomez, A., Hou, B., Toussaint, N., Zimmer, V., Tan, J., et al., 2018. Automatic shadow detection in 2D ultrasound images. In: Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis. Springer, pp. 66–75.

Meng, Q., Matthew, J., Zimmer, V.A., Gomez, A., Lloyd, D.F., Rueckert, D., Kainz, B., 2020b. Mutual information-based disentangled neural networks for classifying unseen categories in different domains: application to fetal ultrasound imaging. IEEE Trans. Med. Imaging 40 (2), 722–734.

Meng, Q., Pawlowski, N., Rueckert, D., Kainz, B., 2019a. Representation disentanglement for multi-task learning with application to fetal ultrasound. In: Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis. Springer, pp. 47–55.

Meng, Q., Rueckert, D., Kainz, B., 2020c. Unsupervised cross-domain image classification by distance metric guided feature alignment. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. Springer, pp. 146–157.

Meng, Q., Sinclair, M., Zimmer, V., Hou, B., Rajchl, M., Toussaint, N., Oktay, O., Schlemper, J., Gomez, A., Housden, J., et al., 2019b. Weakly supervised estimation of shadow confidence maps in fetal ultrasound imaging. IEEE Trans. Med. Imaging 38 (12), 2755–2767.

Meng, Y., Wei, M., Gao, D., Zhao, Y., Yang, X., Huang, X., Zheng, Y., 2020d. CNN-GCN aggregation enabled boundary regression for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 352–362.

Meng, L., Zhao, D., Yang, Z., Wang, B., 2020a. Automatic display of fetal brain planes and automatic measurements of fetal brain parameters by transabdominal three-dimensional ultrasound. J. Clin. Ultrasound 48 (2), 82–88.

Merz, E., Pashaj, S., 2017. Advantages of 3D ultrasound in the assessment of fetal abnormalities. J. Perinat. Med. 45 (6), 643–650.

Miyagi, Y., Hata, T., Bouno, S., Koyanagi, A., Miyake, T., 2021. Recognition of facial expression of fetuses by artificial intelligence (AI). J. Perinat. Med. 49 (5), 596–603.

Moccia, S., Fiorentino, M.C., Frontoni, E., 2021. MaskR^{2}CNN: a distance-field regression version of mask-RCNN for fetal-head delineation in ultrasound images. Int. J. Comput. Assist. Radiol. Surg. 1–8.

Mohamed, F., Siang, C.V., 2019. A survey on 3D ultrasound reconstruction techniques. In: Artificial Intelligence—Applications in Medicine and Biology. IntechOpen London, UK, pp. 73–92.

Montero, A., Bonet-Carne, E., Burgos-Artizzu, X.P., 2021. Generative adversarial networks to improve fetal brain fine-grained plane classification. Sensors 21 (23), 7975.

Morris, S.A., Lopez, K.N., 2021. Deep learning for detecting congenital heart disease in the fetus. Nat. Med. 27 (5), 764–765.

Moser, F., Huang, R., Papageorghiou, A.T., Papież, B.W., Namburete, A.I., 2020. Automated fetal brain extraction from clinical ultrasound volumes using 3D convolutional neural networks. Commun. Comput. Inf. Sci. 151–163.

Namburete, A.I., Xie, W., Noble, J.A., 2017. Robust regression of brain maturation from 3D fetal neurosonography using CRNs. In: Fetal, Infant and Ophthalmic Medical Image Analysis. Springer, pp. 73–80.

Namburete, A.I., Xie, W., Yaqub, M., Zisserman, A., Noble, J.A., 2018. Fully-automated alignment of 3D fetal brain ultrasound to a canonical reference space using multi-task learning. Med. Image Anal. 46, 1–14.

Ni, D., Yang, X., Chen, X., Chin, C.-T., Chen, S., Heng, P.A., Li, S., Qin, J., Wang, T., 2014. Standard plane localization in ultrasound by radial component model and selective search. Ultrasound Med. Biol. 40 (11), 2728–2742.

Nurmaini, S., Rachmatullah, M.N., Sapitri, A.I., Darmawahyuni, A., Jovandy, A., Firdaus, F., Tutuko, B., Passarella, R., 2020. Accurate detection of septal defects with fetal ultrasonography images using deep learning-based multiclass instance segmentation. IEEE Access 8, 196160–196174.

Nurmaini, S., Rachmatullah, M.N., Sapitri, A.I., Darmawahyuni, A., Tutuko, B., Firdaus, F., Partan, R.U., Bernolian, N., 2021. Deep learning-based computer-aided fetal echocardiography: Application to heart standard view segmentation for congenital heart defects detection. Sensors 21 (23), 8007.

Oghli, M.G., Moradi, S., Sirjani, N., Gerami, R., Ghaderi, P., Shabanzadeh, A., Arabi, H., Shiri, I., Zaidi, H., 2020. Automatic measurement of fetal head biometry from ultrasound images using deep neural networks. In: 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference. IEEE, pp. 1–3.

Oghli, M.G., Shabanzadeh, A., Moradi, S., Sirjani, N., Gerami, R., Ghaderi, P., Taheri, M.S., Shiri, I., Arabi, H., Zaidi, H., 2021. Automatic fetal biometry prediction using a novel deep convolutional network architecture. Phys. Med. 88, 127–137.

Oguz, B.U., Wang, J., Yushkevich, N., Pouch, A., Gee, J., Yushkevich, P.A., Schwartz, N., Oguz, I., 2018. Combining deep learning and multi-atlas label fusion for automated placenta segmentation from 3DUS. In: Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis. Springer, pp. 138–148.

Ouahabi, A., Taleb-Ahmed, A., 2021. Deep learning for real-time semantic segmentation: Application in ultrasound imaging. Pattern Recognit. Lett. 144, 27–34.

Patra, A., Huang, W., Noble, J.A., 2017. Learning spatio-temporal aggregation for fetal heart analysis in ultrasound video. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, pp. 276–284.

Patra, A., Noble, J.A., 2019. Multi-anatomy localization in fetal echocardiography videos. In: 2019 IEEE 16th International Symposium on Biomedical Imaging. IEEE, pp. 1761–1764.

Patra, A., Noble, J.A., 2020. Hierarchical class incremental learning of anatomical structures in fetal echocardiography videos. IEEE J. Biomed. Health Inf. 24 (4), 1046–1058.

Perez-Gonzalez, J., Montiel, N.H., Bañuelos, V.M., 2020. Deep learning spatial compounding from multiple fetal head ultrasound acquisitions. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. Springer, pp. 305–314.

Peters, R.J., Iyer, A., Itti, L., Koch, C., 2005. Components of bottom-up gaze allocation in natural images. Vis. Res. 45 (18), 2397–2416.

Płotka, S., Włodarczyk, T., Klasa, A., Lipa, M., Sitek, A., Trzciński, T., 2021. Fetalnet: Multi-task deep learning framework for fetal ultrasound biometric measurements. In: International Conference on Neural Information Processing. Springer, pp. 257–265.

Prieto, J.C., Shah, H., Rosenbaum, A.J., Jiang, X., Musonda, P., Price, J.T., Stringer, E.M., Vwalika, B., Stamilio, D.M., Stringer, J.S., 2021. An automated framework for image classification and segmentation of fetal ultrasound images for gestational age estimation. In: Medical Imaging 2021: Image Processing, Vol. 11596. International Society for Optics and Photonics, 115961N.

Pu, B., Li, K., Li, S., Zhu, N., 2021a. Automatic fetal ultrasound standard plane recognition based on deep learning and IIoT. IEEE Trans. Ind. Inf..

Pu, B., Zhu, N., Li, K., Li, S., 2021b. Fetal cardiac cycle detection in multi-resource echocardiograms using hybrid classification framework. Future Gener. Comput. Syst. 115, 825–836.

Qiao, S., Pang, S., Luo, G., Pan, S., Yu, Z., Chen, T., Lv, Z., 2022. RLDS: An explainable residual learning diagnosis system for fetal congenital heart disease. Future Gener. Comput. Syst. 128, 205–218.

Qiao, D., Zulkernine, F., 2020. Dilated squeeze-and-excitation u-net for fetal ultrasound image segmentation. In: 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology. IEEE, pp. 1–7.

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

Rachmatullah, M., Nurmaini, S., Sapitri, A., Darmawahyuni, A., Tutuko, B., Firdaus, F., 2021. Convolutional neural network for semantic segmentation of fetal echocardiography based on four-chamber view. Bull. Electr. Eng. Inform. 10 (4), 1987–1996.

Rasheed, K., Junejo, F., Malik, A., Saqib, M., 2021. Automated fetal head classification and segmentation using ultrasound video. IEEE Access.

Rawat, V., Jain, A., Shrimali, V., 2018. Automated techniques for the interpretation of fetal abnormalities: a review. Appl. Bionics Biomech. 2018.

Rong, Y., Xiang, D., Zhu, W., Shi, F., Gao, E., Fan, Z., Chen, X., 2019. Deriving external forces via convolutional neural networks for biomedical image segmentation. Biomed. Opt. Express 10 (8), 3800–3814.

Rueda, S., Fathima, S., Knight, C.L., Yaqub, M., Papageorghiou, A.T., Rahmatullah, B., Foi, A., Maggioni, M., Pepe, A., Tohka, J., et al., 2013. Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge. IEEE Trans. Med. Imaging 33 (4), 797–813.

Ryou, H., Yaqub, M., Cavallaro, A., Papageorghiou, A.T., Noble, J.A., 2019. Automated 3D ultrasound image analysis for first trimester assessment of fetal health. Phys. Med. Biol. 64 (18), 185010.

Salomon, L., Alfirevic, Z., Berghella, V., Bilardo, C., Chalouhi, G., Costa, F.D.S., Hernandez-Andrade, E., Malinger, G., Munoz, H., Paladini, D., et al., 2022. ISUOG practice guidelines (updated): performance of the routine mid-trimester fetal ultrasound scan. Ultrasound Obstet. Gynecol.: Off. J. Int. Soc. Ultrasound Obstet. Gynecol..

Salomon, L., Alfirevic, Z., Da Silva Costa, F., Deter, R., Figueras, F., Ghi, T.a., Glanc, P., Khalil, A., Lee, W., Napolitano, R., et al., 2019. ISUOG practice guidelines: ultrasound assessment of fetal biometry and growth. Ultrasound Obstet. Gynecol. 53 (6), 715–723.

Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D., 2019. Attention gated networks: Learning to leverage salient regions in medical images. Med. Image Anal. 53, 197–207.

Schmidt-Richberg, A., Brosch, T., Schadewaldt, N., Klinder, T., Cavallaro, A., Salim, I., Roundhill, D., Papageorghiou, A., Lorenz, C., 2017. Abdomen segmentation in 3D fetal ultrasound using CNN-powered deformable models. In: Fetal, Infant and Ophthalmic Medical Image Analysis. Springer, pp. 52–61.

Schwartz, N., Oguz, I., Wang, J., Pouch, A., Yushkevich, N., Parameshwaran, S., Gee, J., Yushkevich, P., Oguz, B., 2021. Fully automated placental volume quantification from 3DUS for prediction of small-for-gestational-age infants. J. Ultrasound Med..

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Gradcam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 618–626.

Senra, J.C., Yoshizaki, C.T., Doro, G.F., Ruano, R., Gibelli, M.A.B.C., Rodrigues, A.S., Koch, V.H.K., Krebs, V.L.J., Zugaib, M., Francisco, R.P.V., et al., 2020. Kidney impairment in fetal growth restriction: three-dimensional evaluation of volume and vascularization. Prenat. Diagn. 40 (11), 1408–1417.

Sharma, H., Droste, R., Chatelain, P., Drukker, L., Papageorghiou, A.T., Noble, J.A., 2019. Spatio-temporal partitioning and description of full-length routine fetal anomaly ultrasound scans. In: 2019 IEEE 16th International Symposium on Biomedical Imaging. IEEE, pp. 987–990.

Sharma, H., Drukker, L., Chatelain, P., Droste, R., Papageorghiou, A.T., Noble, J.A., 2021. Knowledge representation and learning of operator clinical workflow from full-length routine fetal ultrasound scan videos. Med. Image Anal. 69, 101973.

Sheller, M.J., Edwards, B., Reina, G.A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R.R., et al., 2020. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci. Rep. 10 (1), 1–12.

Shen, Y.-T., Chen, L., Yue, W.-W., Xu, H.-X., 2021. Artificial intelligence in ultrasound. Eur. J. Radiol. 109717.

Sinclair, M., Baumgartner, C.F., Matthew, J., Bai, W., Martinez, J.C., Li, Y., Smith, S., Knight, C.L., Kainz, B., Hajnal, J., et al., 2018. Human-level performance on automatic head biometrics in fetal ultrasound using fully convolutional neural networks. In: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 714–717.

Singh, T., Kudavelly, S.R., Suryanarayana, K.V., 2021a. Deep learning based fetal face detection and visualization in prenatal ultrasound. In: 2021 IEEE 18th International Symposium on Biomedical Imaging. IEEE, pp. 1760–1763.

Singh, V., Sridar, P., Kim, J., Nanan, R., Poornima, N., Priya, S., Reddy, G.S., Chandrasekaran, S., Krishnakumar, R., 2021b. Semantic segmentation of cerebellum in 2D fetal ultrasound brain images using convolutional neural networks. IEEE Access 9, 85864–85873.

Skeika, E.L., Da Luz, M.R., Fernandes, B.J.T., Siqueira, H.V., De Andrade, M.L.S.C., 2020. Convolutional neural network to detect and measure fetal skull circumference in ultrasound imaging. IEEE Access 8, 191519–191529.

Sobhaninia, Z., Emami, A., Karimi, N., Samavi, S., 2020. Localization of fetal head in ultrasound images by multiscale view and deep neural networks. In: 2020 25th International Computer Conference, Computer Society of Iran. IEEE, pp. 1–5.

Sobhaninia, Z., Rafiei, S., Emami, A., Karimi, N., Najarian, K., Samavi, S., Soroushmehr, S.R., 2019. Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning. In: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 6545–6548.

Sofaer, H.R., Hoeting, J.A., Jarnevich, C.S., 2019. The area under the precision-recall curve as a performance metric for rare binary events. Methods Ecol. Evol. 10 (4), 565–577.

Song, C., Gao, T., Wang, H., Sudirman, S., Zhang, W., Zhu, H., 2021. The classification and segmentation of fetal anatomies ultrasound image: A survey. J. Med. Imag. Health Inform. 11 (3), 789–802.

Sree, S.J., Vasanthanayaki, C., 2019. Ultrasound fetal image segmentation techniques: A review. Curr. Med. Imaging 15 (1), 52–60.

Sridar, P., Kumar, A., Quinton, A., Nanan, R., Kim, J., Krishnakumar, R., 2019. Decision fusion-based fetal ultrasound image plane classification using convolutional neural networks. Ultrasound Med. Biol. 45 (5), 1259–1273.

Sun, S., Kwon, J.-Y., Park, Y., Cho, H.C., Hyun, C.M., Seo, J.K., 2021. Complementary network for accurate amniotic fluid segmentation from ultrasound images. IEEE Access 9, 108223–108235.

Sundaresan, V., Bridge, C.P., Ioannou, C., Noble, J.A., 2017. Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging. IEEE, pp. 671–674.

Tan, J., Au, A., Meng, Q., FinesilverSmith, S., Simpson, J., Rueckert, D., Razavi, R., Day, T., Lloyd, D., Kainz, B., 2020. Automated detection of congenital heart disease in fetal ultrasound screening. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. Springer, pp. 243–252.

Tan, J., Au, A., Meng, Q., Kainz, B., 2019. Semi-supervised learning of fetal anatomy from ultrasound. In: Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data. Springer, pp. 157–164.

Torrents-Barrena, J., Monill, N., Piella, G., Gratacós, E., Eixarch, E., Ceresa, M., Ballester, M.A.G., 2021. Assessment of radiomics and deep learning for the segmentation of fetal and maternal anatomy in magnetic resonance imaging and ultrasound. Acad. Radiol. 28 (2), 173–188.

Torrents-Barrena, J., Piella, G., Masoller, N., Gratacós, E., Eixarch, E., Ceresa, M., Ballester, M.Á.G., 2019. Segmentation and classification in MRI and US fetal imaging: recent trends and future prospects. Med. Image Anal. 51, 61–88.

Torrents-Barrena, J., Piella, G., Valenzuela-Alcaraz, B., Gratacos, E., Eixarch, E., Ceresa, M., Ballester, M.A.G., 2020. TTTS-stgan: Stacked generative adversarial networks for TTTS fetal surgery planning based on 3D ultrasound. IEEE Trans. Med. Imaging 39 (11), 3595–3606.

Tsai, P.-Y., Hung, C.-H., Chen, C.-Y., Sun, Y.-N., 2021. Automatic fetal middle sagittal plane detection in ultrasound using generative adversarial network. Diagnostics 11 (1), 21.

Van Sloun, R.J., Cohen, R., Eldar, Y.C., 2019. Deep learning in ultrasound imaging. Proc. IEEE 108 (1), 11–29.

Vaze, S., Namburete, A.I., 2018. Segmentation of fetal adipose tissue using efficient CNNs for portable ultrasound. In: Data Driven Treatment Response Assessment and Preterm, Perinatal, and Paediatric Image Analysis. Springer, pp. 55–65.

Venturini, L., Papageorghiou, A.T., Noble, J.A., Namburete, A.I., 2020. Multi-task CNN for structural semantic segmentation in 3D fetal brain ultrasound. Commun. Comput. Inf. Sci. 164–173.

Wang, L., Guo, D., Wang, G., Zhang, S., 2020. Annotation-efficient learning for medical image segmentation based on noisy pseudo labels and adversarial learning. IEEE Trans. Med. Imaging.

Wang, S., Hua, Y., Cao, Y., Song, T., Xue, Z., Gong, X., Wang, G., Ma, R., Guan, H., 2018. Deep learning based fetal middle cerebral artery segmentation in large-scale ultrasound images. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine. pp. 532–539.

Wang, Q., Ma, Y., Zhao, K., Tian, Y., 2022. A comprehensive survey of loss functions in machine learning. Ann. Data Sci. 9 (2), 187–212.

Wang, X., Yang, X., Dou, H., Li, S., Heng, P.-A., Ni, D., 2019. Joint segmentation and landmark localization of fetal femur in ultrasound volumes. In: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). IEEE, pp. 1–5.

Weerasinghe, N., Lovell, N.H., Welsh, A.W., Stevenson, G., 2020. Multi-parametric fusion of 3D power Doppler ultrasound for fetal kidney segmentation using fully convolutional neural networks. IEEE J. Biomed. Health Inf..

Whitworth, M., Bricker, L., Mullan, C., 2015. Ultrasound for fetal assessment in early pregnancy. Cochrane Database of Systematic Reviews.

Włodarczyk, T., Płotka, S., Rokita, P., Sochacki-Wójcicka, N., Wójcicki, J., Lipa, M., Trzciński, T., 2020. Spontaneous preterm birth prediction using convolutional neural networks. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. Springer, pp. 274–283.

Włodarczyk, T., Płotka, S., Trzciński, T., Rokita, P., Sochacki-Wójcicka, N., Lipa, M., Wójcicki, J., 2019. Estimation of preterm birth markers with U-Net segmentation network. In: Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis. Springer, pp. 95–103.

Wu, L., Cheng, J.-Z., Li, S., Lei, B., Wang, T., Ni, D., 2017a. FUIQA: Fetal ultrasound image quality assessment with deep convolutional networks. IEEE Trans. Cybern. 47 (5), 1336–1349.

Wu, Y., Shen, K., Chen, Z., Wu, J., 2020. Automatic measurement of fetal cavum septum pellucidum from ultrasound images using deep attention network. In: 2020 IEEE International Conference on Image Processing. pp. 2511–2515.

Wu, L., Xin, Y., Li, S., Wang, T., Heng, P.-A., Ni, D., 2017b. Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation. In: IEEE 14th International Symposium on Biomedical Imaging. IEEE, pp. 663–666.

Wyburd, M.K., Hesse, L.S., Aliasi, M., Jenkinson, M., Papageorghiou, A.T., Haak, M.C., Namburete, A.I., 2021. Assessment of regional cortical development through fissure based gestational age estimation in 3D fetal ultrasound. In: Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis. Springer, pp. 242–252.

Wyburd, M.K., Jenkinson, M., Namburete, A.I.L., 2020. Cortical plate segmentation using CNNs in 3D fetal ultrasound. In: Papież, B.W., Namburete, A.I.L., Yaqub, M., Noble, J.A. (Eds.), Medical Image Understanding and Analysis. Springer International Publishing, pp. 56–68.

Xia, T.-H., Tan, M., Li, J.-H., Wang, J.-J., Wu, Q.-Q., Kong, D.-X., 2021. Establish a normal fetal lung gestational age grading model and explore the potential value of deep learning algorithms in fetal lung maturity evaluation. Chin. Med. J. 134 (15), 1828.

Xie, H., Wang, N., He, M., Zhang, L., Cai, H., Xian, J., Lin, M., Zheng, J., Yang, Y., 2020. Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal. Ultrasound Obstet. Gynecol. 56 (4), 579–587.

Xu, Z., Huo, Y., Park, J., Landman, B., Milkowski, A., Grbic, S., Zhou, S., 2018. Less is more: Simultaneous view classification and landmark detection for abdominal ultrasound images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 711–719.

Xu, Y., Lee, L.H., Drukker, L., Yaqub, M., Papageorghiou, A.T., Noble, A.J., 2020c. Simulating realistic fetal neurosonography images with appearance and growth change using cycle-consistent adversarial networks and an evaluation. J. Med. Imaging 7 (5), 057001.

Xu, L., Liu, M., Shen, Z., Wang, H., Liu, X., Wang, X., Wang, S., Li, T., Yu, S., Hou, M., et al., 2020a. DW-Net: A cascaded convolutional neural network for apical four-chamber view segmentation in fetal echocardiography. Comput. Med. Imaging Graph. 80, 101690.

Xu, L., Liu, M., Zhang, J., He, Y., 2020b. Convolutional-neural-network-based approach for segmentation of apical four-chamber view from fetal echocardiography. IEEE Access 8, 80437–80446.

Yang, X., Dou, H., Huang, R., Xue, W., Huang, Y., Qian, J., Zhang, Y., Luo, H., Guo, H., Wang, T., et al., 2021a. Agent with warm start and adaptive dynamic termination for plane localization in 3D ultrasound. IEEE Trans. Med. Imaging.

Yang, X., Huang, Y., Huang, R., Dou, H., Li, R., Qian, J., Huang, X., Shi, W., Chen, C., Zhang, Y., et al., 2021b. Searching collaborative agents for multi-plane localization in 3D ultrasound. Med. Image Anal. 102119.

Yang, X., Shi, W., Dou, H., Qian, J., Wang, Y., Xue, W., Li, S., Ni, D., Heng, P.-A., 2019a. Fetusmap: Fetal pose estimation in 3D ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 281–289.

Yang, X., Wang, X., Wang, Y., Dou, H., Li, S., Wen, H., Lin, Y., Heng, P.-A., Ni, D., 2020. Hybrid attention for automatic segmentation of whole fetal head in prenatal ultrasound volumes. Comput. Methods Programs Biomed. 194, 105519.

Yang, X., Yu, L., Li, S., Wang, X., Wang, N., Qin, J., Ni, D., Heng, P.-A., 2017. Towards automatic semantic segmentation in volumetric ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 711–719.

Yang, X., Yu, L., Li, S., Wen, H., Luo, D., Bian, C., Qin, J., Ni, D., Heng, P.-A., 2019b. Towards automated semantic segmentation in prenatal volumetric ultrasound. IEEE Trans. Med. Imaging 38, 180–193.

Yaqub, M., Kelly, B., Papageorghiou, A.T., Noble, J.A., 2017. A deep learning solution for automatic fetal neurosonographic diagnostic plane verification using clinical standard constraints. Ultrasound Med. Biol. 43 (12), 2925–2933.

Yu, L., Guo, Y., Wang, Y., Yu, J., Chen, P., 2017a. Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks. IEEE Trans. Biomed. Eng. 64 (8), 1886–1895. http://dx.doi.org/10.1109/TBME.2016.2628401.

Yu, Z., Tan, E.-L., Ni, D., Qin, J., Chen, S., Li, S., Lei, B., Wang, T., 2017b. A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition. IEEE J. Biomed. Health Inf. 22 (3), 874–885.

Zaffino, P., Moccia, S., De Momi, E., Spadea, M.F., 2020. A review on advances in intra-operative imaging for surgery and therapy: Imagining the operating room of the future. Ann. Biomed. Eng. 1–21.

Zeng, Y., Tsui, P.-H., Wu, W., Zhou, Z., Wu, S., 2021. Fetal ultrasound image segmentation for automatic head circumference biometry using deeply supervised attention-gated V-Net. J. Digit. Imaging 34 (1), 134–148.

Zhang, B., Liu, H., Luo, H., Li, K., 2021. Automatic quality assessment for 2D fetal sonographic standard plane based on multitask learning. Medicine 100 (4).

Zhang, J., Petitjean, C., Lopez, P., Ainouz, S., 2020a. Direct estimation of fetal head circumference from ultrasound images based on regression CNN. In: Medical Imaging with Deep Learning. PMLR, pp. 914–922.

Zhang, J., Petitjean, C., Yger, F., Ainouz, S., 2020b. Explainability for regression CNN in fetal head circumference estimation from ultrasound images. In: Interpretable and Annotation-Efficient Learning for Medical Image Computing. Springer, pp. 73–82.

Zhang, L., Portenier, T., Paulus, C., Goksel, O., 2020c. Deep image translation for enhancing simulated ultrasound images. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. Springer, pp. 85–94.

Zhang, L., Zhang, J., Li, Z., Song, Y., 2020d. A multiple-channel and atrous convolution network for ultrasound image segmentation. Med. Phys. 47 (12), 6270–6285.

Zhao, C., Droste, R., Drukker, L., Papageorghiou, A.T., Noble, J.A., 2021. Visual-assisted probe movement guidance for obstetric ultrasound scanning using landmark retrieval. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 670–679.

Zhu, F., Liu, M., Wang, F., Qiu, D., Li, R., Dai, C., 2021. Automatic measurement of fetal femur length in ultrasound images: a comparison of random forest regression model and SegNet. Math. Biosci. Eng. 18 (6), 7790–7805.

Zimmer, V.A., Gomez, A., Skelton, E., Ghavami, N., Wright, R., Li, L., Matthew, J., Hajnal, J.V., Schnabel, J.A., 2020. A multi-task approach using positional information for ultrasound placenta segmentation. In: Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. Springer, pp. 264–273.

Zimmer, V.A., Gomez, A., Skelton, E., Toussaint, N., Zhang, T., Khanal, B., Wright, R., Noh, Y., Ho, A., Matthew, J., et al., 2019. Towards whole placenta segmentation at late gestation using multi-view ultrasound images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 628–636.