Proc. of the International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)
16-18 November 2022, Maldives

Segmentation of Fetal 2D Images with Deep Learning: A Review

Pedro Rodrigues CEDRI
Instituto Politécnico de Bragança
Bragança, Portugal
pjsr@ipb.pt

M. Anees ur Rehman
University of Engineering and Technology
Taxila, Pakistan

Getúlio Igrejas CEDRI
Instituto Politécnico de Bragança
Bragança, Portugal
igrejas@ipb.pt

***Abstract*—Image segmentation plays a vital role in providing sustainable medical care in this evolving biomedical image processing technology. Nowadays, it is considered one of the most important research directions in the computer vision field. Since the last decade, deep learning-based medical image processing has become a research hotspot due to its exceptional performance. In this paper, we present a review of different deep learning techniques used to segment fetal 2D images. First, we explain the basic ideas of each approach and then thoroughly investigate the methods used for the segmentation of fetal images. Secondly, the results and accuracy of different approaches are also discussed. The dataset details used for assessing the performance of the respective method are also documented. Based on the review studies, the challenges and future work are also pointed out at the end. As a result, it is shown that deep learning techniques are very effective in the segmentation of fetal 2D images.**

Keywords—Image Segmentation, Fetal images, Deep Learning techniques, CNN

I. INTRODUCTION

Medical image segmentation refers to the detection of boundaries or extraction of desired objects (organs) from 2D or 3D images, which can be done manually or automatically. Segmentation of medical images plays a vital role in computer-aided diagnosis systems and covers a wide range of applications [1]. Since the last two decades, the field of medical imaging has progressed at a rapid pace with the integration of artificial intelligence-based processing models [2]. Ultrasound computed tomography (CT), X-ray, Microscopy, Positron Emission Tomography (PET), Dermoscopy, and Magnetic Resonance Imaging (MRI) are some of the well-known modalities used for obtaining medical images [3]. Image segmentation is considered one of the most essential techniques in the medical field because it helps in locating the Region of Interest (ROI) through a semi or fully automatic process. In medical images, the segmentation process divides the image into different parts based on certain descriptions such as border detection, mass detection, or tumor detection. For professionally extracting the ROI from the background, often clustering approach is employed along with the segmentation [4]. Ultrasound (US) images have been commonly utilized for pregnant women during different stages of pregnancy. The examination of fetal development is done by evaluating different kinds of biological parameters such as Femur Length (FL), Biparietal Diameter (BPD), Abdominal Circumference (AC), Crown-rump length (CRL), and Head Circumference (HC). The assessment of fetal Head Circumference (HC) is the most important aspect in the fetal examination, which is measured in a specific cross-section of the fetal head [5-7]. At present, the fetal images are mostly examined through traditional methods such as deformable models [8], the difference of Gaussians [9], texture Maps [10], Morphological Operators [11], and Multilevel thresholding [12]. These techniques are tedious and time-consuming at the same time, which often causes inter and intra-operator differences.

Moreover, in most of these approaches, it is assumed that the fetal head contour is elliptical. Though US images are mostly used for fetal examination, they lack in terms of Speckle Noise, artifacts, low signal-to-noise ratio, and attenuation. To counter these problems, various Machine Learning approaches are also proposed by the researchers. Lu et al. [13] present an iterative Hough transform technique for the inspection of incomplete ellipses in fetal images with intense noise. This study shows promising results in terms of inter-run variation, which was 2.08% for HC and 0.84% for the BPD. The major limitation of this study is that this Machine Learning-based model couldn't detect the fetal head in low contrast ultrasound images. Another similar study by Li et al. [14] proposed a Machine Learning-based random forest model that utilized prior knowledge to locate the region of interest and examine the fetal head with phase symmetry. The evaluation of the presented model on 145 HC images shows an overall measurement error of 1.7mm and outperformed the traditional fetal image analysis methods. But in the case of late pregnancy, this model cannot fit the fetal skull that is partially missing on ultrasound images.

Though Machine Learning based models significantly enhanced the segmentation process of fetal images, they showed significant limitations in clinical testing. To counter these limitations, deep learning-based models were introduced by the researchers for accurate processing and analysis of fetal 2D images. The application of Convolutional Neural Networks (CNN) in medical imaging is considered a major breakthrough, especially in terms of segmentation [15]. Three-dimensional V-Net, Fully Convolutional Network (FCNs), and U-Net are some of the prominent architectures of CNNs. Lingyun Wu et al. [16]

978-1-6654-7095-7/22/$31.00 ©2022 European Union

proposed a cascaded convolutional network for prenatal ultrasound image segmentation automatically. This deep learning model used the CNN algorithm to extract the features from the multiple visual scales. For enhancing the local spatial consistency of the prediction map, researchers employed FCN core with an auto-Context scheme. First, they customized the FCN-8s network to predict the anatomy regions efficiently. Next, researchers implement customized FCN into an Auto-Context model to enhance the prediction map successively. Lastly, they employed a summation operator instead of parallelization to assess the prediction map on the auto-context scheme. Similar to this study, a variety of deep learning techniques have been developed by researchers in the last two decades.

In this paper, we present a detailed literature survey of different deep learning methods used to segment fetal 2D images. We thoroughly investigate each research study and provide a core concept in the next section. The rest of the paper is presented as follows: Literature survey, Conclusion, and Future Work.

II. Literature Survey

Ultrasound imaging is considered the safest and non-invasive method for examining internal body organs. Compared to the other imaging techniques like MRI and CT scans, US imaging is prevalent, portable, and cheaper at the same time. Due to this reason, US imaging is widely used to treat medical conditions. Baidaa Al-Bander et al. [17] proposed a deep learning module to segment the Fetal Model by computing different ultrasonic images. The proposed model can easily be classified into two distinct categories, which are localization and segmentation. Combined object localization can act as a fetal head boundary, whereas the segmentation can be directly used for the deep learning module. The proposed network is a fully convolutional network wherein the accuracy of the network will be further enhanced by employing the combined object localization. A least-square ellipse model is used for the computation of segmented fetal heads. The integrated model was trained on a 2D ultrasonic images dataset.

The proposed module of the investigators can be classified into two different categories. The first part of the model is the segmentation of the fetal head with the base structure of Regional Convolutional Neural (R-CNN). The second part of the module is the fitting model, which was based on the least square fitting method. R-CNN was originally designed for employing segmentation and localization in one architecture. The delved module can be subdivided into four major parts as follows

A. Extraction of features is the first step in the proposed model. This step consists of pooling and convolutional layers, which serve as the major contributor to the extraction of segmentation. 2D ultrasonic images were resized as 512-512-3 for compatibility with the module. The model was not trained from scratch, but it employed Resnet50, for which the first 50 layers of the module were selected. This led to a faster and more responsive R-CNN.

B. Object localization was then computed and presented by the fetal head, for which the main computational factor was R-CNN. The model was then able to generate several ROIs with the help of bounding boxes prediction. Object localization was achieved by employing RPN (Region Proposal Network). The RPN ran the scans from the feature maps and resulted in outputs of specimen anchors. The anchors were then assessed by classifier and regressor for repetitive regions. Two outputs were provided by the RPN, namely as background and foreground and the refinement of anchor box location. Later, top anchors were selected as they were more likely to contain the ROI.

C. ROI align module was employed to adjust the dimension of the ROIs as acquired by RPN. ROI Align is a technique that uses bilinear interpolation to produce the feature maps of fixed sizes. The classifier then computes whether an ROI is positive or negative.

D. Positive regions or fetal head regions are fed to Mask R-CNN. It is a fully coevolutionary network that generates masks with localized ROI.

The proposed module was tested by employing the Dice coefficient. Investigators report that a result of 97.73±1.32 was acquired with this model.

Hariharan Ravishankar et al. [18] proposed a hybrid approach to detecting abdominal contour. The proposed model is a combination of deep learning automatic detection and traditional texture analysis. A learning module was employed for the segmentation of ROI and abdominal boundaries. A significant improvement was observed in segmentation and ROI detection when the model was combined with GBM (gradient boosting machine) and HOG (Histogram of oriented gradient). The proposed model runs the entire dataset with an ML classifier and the retention of the positive candidate, which possesses the positive label class. For building the classifier, users swept the field of view (FOV), which enabled them to generate multiple patches which are overlapping. However, the size of the path must be aligned with the defined abdominal circumference. The ROIs were then resized to a standard size. ROIs were then differentiated into two different classes with attributes as overlap, ground truth, and ground-truth ellipse. A CNN network was trained for the computation, which consisted of two max-pooling, two convolutional, and a single softmax layer. Both convolutional layers were set to output at 24 with kernel sized as 12-12 and 6-6 with two different layers. Whereas the pool strides the max for the pool layers was 2-2. Rectified Linear Units or ReLU's were employed as an activation function for the model. For the part of traditional statistical methods, the GBM and HOG were compared at length by the investigators. They present that the HOG has outperformed all other conventional modules by capturing the measuring orientation and image gradients. These details enabled the HOG modules to perform efficiently even under unlikely and non-uniform conditions. However, GBM presented the best results when it comes to feature vector dimensions. GBM was employed by the investigators in the proposed model with the parameters as shrinkage factor set to 0.5, tree depth to 2, and iterations to 200. For detecting the region of the fetal abdomen, the values with the maximum likelihood of positive labels were selected within each image. GBM classifier then provided whether any value bears positive value or not.

The designed model was trained with 70 different images and later was tested with another set of 70 images. A dice similarity coefficient of 0.90 was obtained, which is a good indication of overlap with the ground truth. The mean computed gestational age difference between the computed and actual is within 90% for two weeks and can drop to 70%

for one week. The investigator concludes with the proposed model that Convolutional neural networks are far better at performance when compared with traditional methods.

Zahra Sobhaninia et al. [19] proposed a neural network-based model for automatic segmentation of fetal Ultrasound images. The module can predict the High Capacity (HC) ellipse with a higher degree of accuracy. The model is also an assurance of minimization for the cost function consisting of Mean Squared Errors (MSE) parameters as well as dice score of segmentation. A simple illustration of the proposed model is as follows. The proposed model was designed based on Multi-Layered Traffic Network (MLTN) or link-net architecture with the inclusion of multiple inputs. The architecture enabled the module to work as an end-to-end multi-task network system. The module was majorly classified into two components, namely a network segmentation and an ellipse tuner. Two functions were employed as boundary functions which assured the improvement in the training process for the entire network. The training phase employed different entry points. The training module was trained by back-propagation of the loss gradients combined. The segmentation network proposed by the investigators is a modified version of the Link-Net network, which possesses the capabilities of multiple inputs. The data acquired by MLTN is 2D ultrasonic images with segregation of three different scales, which are then computed and fed to different layers accordingly. Pooling and convolutional layers are integrated with the first part of the network, which serves an encoding mechanism usually termed as ResBlock. The concatenation of the first and second feature maps provided the flexibility of a multi-structured module. The later part of the network is the decoder block which ensures the upsampling and building of final segmentation outputs. Different sorts of skip-connection are present for effective correspondence between encoders and decoders. Skip connections assure that feature-map details are well preserved, ultimately leading to a precise segmentation of boundaries. Radiologists often encounter elliptical images for assessment. Hence, an ellipse tuner was integrated with the segmentation network for accurate data computation. The ellipse parameters were computed for the final assessment of the fetal head. For this assessment, the rich feature map module was integrated into the middle of the segmentation network. The ellipse parameters were calculated between the encoder and decoder of the segmentation network. The extracted features were then fed to three fully connected (FC) layers for tuning. The outcome of the tuning is five different outputs which serve as the basis for the estimation of ellipse parameters as well as fetus head location and shape. Investigators suggest that the proposed FC layers significantly contribute to segmentation and improve the model's overall accuracy.

Lingyun Wu et al.[20] proposed a computational scheme as computerized fetal US image quality assessment (FUIQA). Investigators aimed at improving the implementation of US image quality control and the assistance in obstetric examinations. The proposed module is designed by employing two convolutional neural network models, which were terms as L-CNN and C-CNN for the case of segregation. The first convolutional network, or L-CNN, was implemented to find out the region of interest (ROI) of the fetal abdominal region in any given US picture. The L-CNN layers, specifically C and M were initialized by using AlexNet. However, the F layer of the module was initiated with the help of a Gaussian randomizer. The computational output of this module is then directly fed to the C-CNN. C-CNN network evaluated the image quality by assessing different attributes of the image such as the quality of depiction for the features and key structures. The key structures assessed by C-CNN are the umbilical vein and stomach bubble. To further enhance the performance of the designed model, specifically L-CNN, the input sources of neural networks were augmented. The same was carried out by using the local phase feature and backed by original US data. Heterogeneous sources of inputs provided more accuracy in the prediction and enhanced computational analysis. With heterogeneous and multiple inputs of the data L-CNN was able to predict the symmetric and asymmetric features from the images. This empowered the L-CNN module to detect the ROI for fetal abdominal localization. Later the C-CNN analyzed the identified ROI by employing different classification schemes. A four-class learning framework was integrated for C-CNN architecture. This enabled the module to differentiate between SB, UV, and SP and the planes with the coexistence of SB, GB, and SP. C and M layers share a similar structure to the L-CNN module. However, the F-layer was programmed to a different structure for reasoning purposes. The F-layers were set randomly from the Gaussian distribution. Additionally, dynamic adaptation and dropout strategies for the learning rate of L-CNN were also integrated with C-CNN. The topmost F-layer was connected with the softmax, which presented the results of classification outputs. C-CNN rates the effectiveness of ROI for any US Image with a different score. For example, it will score 2 if the UV in the image is not well assessed or depicted. The proposed model for assessment was compared by subjective quality evaluation from three different doctors. Investigators proposed that the computational assessment method can lead to a higher degree of accuracy.

Guotai Wang et al. [21] proposed a different approach toward segmentation. The proposed model suggests integrating an interactive framework featuring a deep learning module. The model also employs CNN within the bounding box and a pipeline of segmentation designed based on scribble. Image-specific fine-tuning has also been implemented, ensuring the flexibility of adaptation to any given image. The adaptation of the model can be specified by one of two learning methods which are supervised and unsupervised learning. For the sake of ultra-fine-tuning of the proposed model, a weighted loss function is also suggested, which can account for the network and interaction-based uncertainty and perform necessary addendums. A descriptive illustration of the model follows.

The model consists of CNN with a bounding box to assess the previously unseen objects. In this case, the model takes content input from the bounding box and returns a binary segmentation for any particular instance. For testing proceedings, user input is provided in a bounding box which is then extracted by BIFSeg. The same is then assessed by CNN, which is pre-trained by a learning dataset to acquire the preliminary segmentation. CNN's are commonly trained for capturing attributes such as contrast, hyperintensity, and saliency for any given object, which results in generalization of the objects which are not visually accessible. Investigators propose employing supervised or unsupervised modules as it is likely that the model may present a mismatch among features. Hence the fine-tuning is integrated to increase the overall accuracy of the model by leveraging specific features. The proposed framework presents flexibility as it can handle both 2D and 3D images by certain network structure assumptions. BIFseg mainly focuses on objects which have not been seen earlier and adapts the CNN to different images by serving user interactions as a computational input. The P-Net framework has been proposed for the bounding box segmentation of images. The proposed network, by making use of dilated convolution, preserves the resolution. A set of six different blocks is present, which are responsible for capturing corresponding attributes. The first block possesses dilation parameters and thus identifies features at multiple scales. The rest of the blocks' work are concatenated, and the output of the same is fed to the final sixth block, which acts as a classifier. However, for the processing of 3-D images, an extended P-Net module is employed. The network possesses a receptive field of 85-85-9, which is anisotropic by nature. Anisotropic resonance is a commonly used technique in MRI. The classifier was fine-tuned with already known features. The model's novelty is that it assesses all the pixels of the image equally. A weighted approach is also proposed for the pixels. Wherein, user-provided scribbles possess a higher degree of confidence and ultimately lead to a higher impact. The investigators utilized the Caffel library to implement P-Net and Pc-Net modules. Automatic generation of the bounding box was permitted during the training phase with the input of the ground truth label and a pixel margin of 0 to 10. The investigators used a cross-entropy loss function decent with a momentum value reaching 0.9.

Seyed Sadegh Mohseni Salehi et al. [22] proposed a fully automated segmentation module that was based on a convolutional neural network. The proposed module was able to compute the fetal brain data in real-time by using 2D MRI slices. The proposed model was structured upon 2D U-net and an auto context. U-net style architecture is mainly composed of two different paths, which are expanding and contracting. The contracting path is powered by a padded 3-3 convolution with a later integration of RelU layers of non-linear nature. The pooling of each convolution is achieved by using 2*2 along with stride 2. The output of downsampling acquired by pooling can result in doubled features. Unlike the contracting path, expanding path mainly focuses on upsampling operation wherein two convolutional operations are also present. The features acquired from the contracting path then undergo concatenation with corresponding feature maps as acquired from the expanding path. The final layer of 1-1 is applied to acquire the linear output from the program. The linear output is a feature map that determines brain tissue or non-tissue by employing the depth of features. The output layer from the network is of two planes, each corresponding to a certain class. Softmax was integrated, which helped information of the loss. The reshaping of the output was carried out with the application of the cross-entropy principle. Balancing of the training samples for different classes was computed and corrected by the mean of each class. It is worth mentioning that the weight of the classes is inversely proportional to that of appearance. Thus, a higher number of appearances led to lower weights. ADAM optimizer was used for the sake of cost optimization by stating an initial learning rate of 0.0001 with a multiplication factor of 0.9 for each 2K steps. It is worth noting that the segmentation on the brain sections was completely independent, which led to efficient brain structures. The proposed framework was trained by employing 250 different stacks of training images dataset. The model was assessed for medium to high degree complex cases. Seventeen stacks of the image were normal fetal brain images, whereas eighteen stacks of complex images were also computed. The complex stack had a severely abnormal shape of brain, noise, and extreme motion making it more difficult to render. The assessment on these stacks presented a dice coefficient of 96.52 % for normal and 78.83% for complex brain images, which outperformed all the other established methods.

Wu lingyun et al. [16] presented a model of fully automated segmentation consisting of a cascading framework. The model can be majorly divided into two different portions. Firstly, a fully convolutional network was integrated for the detection of features from different visual images. Secondly, for the differentiation of anatomy, a dense prediction map was used. To increase the efficiency of the proposed model, investigators used an auto context scheme that was implanted in a fully convolutional network core. The major improvement was achieved by using summation instead of parallelization with auto context. A simple illustrative explanation of the proposed model follows.

A fully convolutional network was trained to compute the fetal anatomy region. A fully convolutional network also highlights the areas of dense boundary maps. Acquired boundary maps are then fed to different levels of auto context. For each level of auto context, the input is a fetal image with feedback from the prediction map, which is acquired or computed from a previous level. For the start, the initial-boundary map level is set to zero, which was considered void. The prediction map is then gradually undergoing refinement by different levels of auto context. The output from the final context level is termed a segmentation mask. As per the investigators, the most successful fully convolutional network structure is the one that consists of a Vgg 16- layer net and was employed for this study. This fully convolutional network model does not take into consideration the final classification layer, and all the other layers are replaced with convolutional layers. The 6th and 7th layers of this fully convolutional network type are capable of generating 4096 feature maps for the segmentation of 21 different classes. Since the proposed model is only a two-class task, so the 6th and 7th layer of the fully convolutional network was omitted to avoid complexities in the model. This also helped prevent the possible overfitting of the data within the model. 1-1 convolutional layers were appended, which generated two different score maps for the fetal anatomy and background for each course output. A deconvolution layer followed the model, which was used to upsample the coarse maps acquired, which led to the computation of pixel-wise outputs. The custom fully convolutional network designed was trained on the model images from the U.S to avoid the overfitting problem when used for a relatively smaller dataset. The cross-entropy function was applied for the training proceedings of the designed model.

S. Jayanthi Sree et al. [23] proposed an automated segmentation based on fuzzy connectedness to evaluate fetal ultrasound images or scans. The proposed segmentation model adapts to the fuzzy nature of the ultrasonic scans. However, the proposed model is not entirely an automated technique and can be termed semi-automatic. The employed method makes use of different and multiple see points to compute the anatomical segmentation of any fetal specimen. The model is designed on the basis that an inherent fuzziness is always present in an ultrasonic image. The model computes the segmentation based on region. The adjacency of pixels for any given image by intensity, spatial, and strength is collectively called fuzzy connectedness. The proposed model consists of two major algorithm modules, which are local and global adjacency, also known as affinity and connectedness, respectively. Take any two pixels as x and y, linking between these two adjacent pixels is called affinity or local fuzzy relation. The affinity function for any given instance is purely dependent on homogeneity of the region as well as intensity difference in values against the intensity value for ROI. However, the global adjacency, which is also called fuzzy connectedness, is evaluated by using affinities. For any given instance, the most elongated path for any x and y pixels can present the connectedness between adjacent pixels. All the paths in corresponding or adjacent pixels form the sequence from x to y. The least affinity for any given pair of pixels can be termed as the strength of the path along the path which is to be computed. Integration of local and affinity adjacency models can help in the formation of a connectivity map. Later on, thresholding is applied to the connectivity maps, which results in the estimation of a region of interest. Several qualitative and quantitative matrixes have been proposed in this study to examine the output of algorithms such as Variation of Information (VOI), Probability Rand Index (PRI), and Global Consistency Error (GCE). The proposed model was tested with all of these indices and showed a relatively better result. Conclusively, the proposed algorithmic module was assessed by using 300 different sets of ultrasonic images for efficacy which returned that the model was 90% effective.

Jinpeng Li et al. [24] proposed an extensive automated model for the problem of fetal brain extraction in 2D among MRI slices. The proposed model was completely powered by fully convolutional networks. Two different FCNs were employed: shallow FCN and the other as Deep Multiple scale FCN (M-FCN). Shallow FCN was able to efficiently and quickly locate the fetal brain. Later, processed for the computation of the region of interest for the brain. The acquired information was then computed by M-FCN, which fine-tuned the segmentation and was able to produce the final brain mask by employing all the potential of multi-scale computation and information processing. Dilated convolutional layers were employed to enhance the field of view and feature maps in both shallow and deep scale FCN. The two-step framework was capable of handling localization and segmentation by making use of the two integrated FCN's. The rapid location of the fetal brain and extraction of the ROI from any 2D images were part of the localization. As for the segmentation, Deep multi-learning FCN refines the segmentation from the ROI acquired by shallow FCN. Dilated convolutional layers and residual learning blocks were implanted within the FCN architecture to acquire more tuned results. The proposed framework was then assessed for efficacy by using 88 2D fetal MRI images. As reported by the investigators, the method outperformed all other established models not only in the area of localization but also in segmentation. Estimation of the location was 100% accurate with the help of shallow FCN in the model. As for the part of segmentation, an efficacy of 0.958 dice score is listed with the sensitivity of 0.950 and a whooping precision score of 0.968 with the employed test dataset. The model was able to produce these incredible results by only 6s as processing time. The system used for the computation was with a configuration of Core I7 6th Generation processor as reported by the investigators.

Yan Zeng et al. [25] proposed a fully automated deep learning module capable of detecting and estimating the segmentation in ultrasonic images. The proposed model consisted of HC biometry, assisted by DAG (deeply supervised attention-Gated) V-Net. The incorporation of HC biometry adds the flexibility of attention addition to any given V-Net modules. A multi-scale loss function was also suggested and integrated as an effective aid in deep supervision of the overall module. The proposed DAG V-Net model was trained by the HC 18 dataset, which was further expanded with the help of augmentation. The resulting model of DAG V-Net was able to segment the fetal head from any 2D ultrasonic image of the specimen. The segmentation was followed by different computations such as processing of morphology, edge detection, and ellipse fitting for effective and efficient results. The resultant ellipses were then directly measured by HC biometrics. Firstly, the investigators enhanced the DAG V-Net module with an augmentation-assisted dataset. The testing set images did not result in a standard measurable ellipse. To eradicate this problem, investigators integrated processing of morphology which led to the elimination of all irrespective circular structures and aided in the elimination of noise for output images. Later on, morphologically corrected or assorted images were fed for edge detection. Subsequently, the least square method was integrated to define the boundary of the detected fetal ellipse. The final calculation of the HC was carried out by using the resolution of pixels for each specimen ultrasound image. This resulted in accurate determination of the fetal head corresponding to the ellipse's rotation angle and center points. The proposed model was assessed for efficacy by employing the HC18 dataset, wherein the 2D images reach 355. Four different performance coefficients were used to measure the performance of the proposed model, which are ADF (HC absolute difference), DSC (Dice similarity coefficient), HD (Hausdorff distance), and DF (HC difference). As reported by the investigators, the experimental results provided the efficacy of the model as follows; Dice similarity coefficient: 97.93%, HC absolute difference: 1.77 ± 1.69 mm, HC difference: 0.09 ± 2.45 mm, Hausdorff distance: 1.29 ± 0.79 mm. Investigators suggested that by integrating attention and deep supervision, the presented model could supersede the conventional U-Net or V-Net models. This resulted in overall efficient segmentation when compared to other proposed published models.

The four standard heart views are naturally segmented by the Siti Nuramani et. al. [26], proposed deep learning-based computer-aided fetal cardiac ultrasound tests with an instance segmentation approach, which concurrently detects the abnormality and segments the four standard heart views. With 1149 fetal heart images, they performed many tests to predict 24 objects, including four conventional fetal heart view shapes, 17 heart chamber objects in each view, and three congenital heart disease scenarios. The suggested model demonstrated acceptable performance for conventional views segmentation, with an intersection over union of 79.97% and a Dice coefficient similarity of 89.70%. With mean average precision of 98.30% for intra-patient variance and 82.42% for inter-patient variation, it also performed well in the detection of coronary heart diseases. Automated classification and detection methods could significantly increase the rate of fetal heart disease identification.

Sreelakshmy R. et. al. [27], presented a novel deep learning (DL) method for automatically segmenting the fetal cerebellum from 2-dimensional US brain images. ReU-Net, a semantic segmentation network made specifically for the fetal cerebellum, is presented by them. Additionally, to separate the cerebellum from the noisy US data, they employed U-Net as a baseline model with the addition of residual blocks and Wiener filter over the last two layers. They collected 590 photos for training and 150 for testing, and also used a 5-fold cross-assessment approach. In terms of Dice Score Coefficient (DSC), F1-score, Hausdorff Distance (HD), accuracy, recall, and precision, the ReU-Net received scores of 91%, 92%, 25.42, 98%, 92%, and 94%, respectively. The proposed method performs significantly better than the existing U-Net based solutions. The proposed method can be used to provide high bandwidth imaging methods in fetal US photos for medical research as well as biometric analysis on a larger scale.

These are some of the latest deep learning methods presented by researchers in recent years. Table 1 shows the summary of the aforementioned studies.

TABLE I. SUMMARY OF AFOREMENTIONED DEEP LEARNING MODELS FOR FETAL IMAGE SEGMENTATION.

Author/ Reference	Modality	Method	Dataset	Performance Evaluation
Baidaa Al-Bander et al. [17]	Ultrasound Images	Object Localisation	999 two-dimensional ultrasound images from the database of DORUMC	DICE: 97.73 % Harsdorf distance (mm) of 1.39
Hariharan Ravishankar et al. [18]	Ultrasound Images	Hybrid approach (CNNs + traditional texture analysis	Custom 140 images dataset	DICE similarity coefficient of 0.90
Zahra Sobhaninia et al. [19]	Ultrasound Images	Multi-task Deep Convolutional Neural Network	999 two-dimensional ultrasound images from the database of DORUMC	DSC: 96.84 % DF (mm): 1.13 ADF (mm): 2.12 HD (mm): 1.72
Lingyun Wu et al.[20]	Ultrasound Images	L-CNN and C-CNN	8072 Ultrasound Images	ROI (Acc) : 0.928 UV (Acc): 0.969
Guotai Wang et al. [21]	MRI and FLAIR images	CNN with image-specific fine-tuning	274 Images from BRATS	DICE: 87.49% T *m (s):* 4.72
Seyed Sadegh Mohseni Salehi et al. [22]	MRI	Deep Fully Convolutional Neural Network (2D U-Net)	285 MRI images	DICE: 96.52% Sensitivity: 94.60% Specificity: 99.92%
Wu lingyun et al. [16]	Ultrasound Images	Cascaded Framework	900 images of the fetal head and 688 images of the abdomen	DICE: 0.9843 Adb: 2.04 Jaccard: 0.9690
S. Jayanthi Sree et al. [23]	Ultrasound Images	Fuzzy Connectedness Algorithm	300 ultrasound fetal images	JI: 0.6853 DC: 0.9023 GCE: 0.0256
Jinpeng Li et al. [24]	MRI	Shallow FCN and an extra deep multi-scale FCN (M-FCN)	30 MRI images	DICE: 0.958 Recall: 0.950 Precision: 0.968
Yan Zeng et al. [25]	Ultrasound Images	Deeply Supervised Attention-Gated V-Net	551 US images from HC18 Challenge	DSC: 97.93% DF (mm): 0.09 HD (mm): 1.27
Siti Nuramani et. al. [26]	Ultrasound Images	Instance Segmentation	1149 fetal heart images	DSC: 89.70% Precision (intra-patient): 98.30% Precision (inter-patient): 82.42%

R. Sreelakshmy et. al. [27]	Ultrasound Images	ReU-Net	590 photos for training and 150 for testing	DSC: 91% F1-Score: 92% HD: 25.42 Accuracy: 98% Recall: 92% Precision: 94%

III. Results and Discussion

Ultrasonic imaging and Magnetic resonance imaging are some of the most commonly employed techniques for disease evaluation. Imaging profiles acquired by using sound or MRI can help the doctors to find out about the disease. However, one of the major problems encountered with MRI or ultrasonic imaging is that the output images are often fuzzy and not clear with plenty of noise. Even though the latest development in technology has offered better rendering, there is still a long way to go. It requires both an expert machine operator and the most commonly assessed parameters for expecting women. The same is carried out by using ultrasonic imaging, which provides the location as well as the size of the fetal head. Lately, several deep learning algorithms have been trained and employed for the recognition of fetal heads.

In this research, many modules with different deep learning configurations have been discussed at length. The study aims to encapsulate the latest data on the subject with the inclusion of semi-automatic as well as fully automated algorithms. For all the discussed deep learning models, effective segmentation is one of the most vital factors for computation. Effective segmentation and accurate evaluation of ROI can lead to more accurate and proficient results. Some of the common highlights follow. A model including two fully convolutional networks presented some promising results. One FCN was allocated for the localization, while the other served the segmentation. The model achieved an accuracy of 97.73±1.32 when assessed by the Dice Coefficient method.

Similarly, a different model was proposed by an investigator with the inclusion of GBM and HOG along with an FCN. The model was able to achieve a 0.90 dice coefficient. Another researcher proposed the FUIQA (computerized fetal US image quality assessment) based on L-CNN and C-CNN. The model was subjectively evaluated by experts in the field and presented a considerable output. A different module was proposed for the integration of V and U-Networks with the inclusion of the bounding box to the FCN. The model's output was fine-tuned by entropy functions which lead to better and optimum results. A model with multiple convolutional layers was proposed containing padding, pooling, and Softmax layers with a dice coefficient efficacy of 96.52%. Integration of FCN with a dense mapping strategy led to the development of different modules that could eradicate the problem of overfitting. A semi-automated model consisting of fuzzy connectedness was also proposed, which utilizes multiple see points for the computation. The model's efficacy with 300 test images was reported to be 90%. Another proposal suggests the integration of HC biometry with DAG V-Net, which resulted in an overall efficacy of 97.93% when tested by Dice Coefficient. Lastly, a dilated convolutional-based model is also reported, with efficacy of 0.958 for dice coefficient. It is worth mentioning that neither of the listed models is purely accurate for all the problems associated with the segmentation and evaluation of ROI. It must be understood that a model getting a good result in one area may well possibly lag in another, e.g., none of the listed models maybe be able to provide the elimination of noise as well as 99 to 99.5 dice efficient for segmentation.

IV. Conclusion and Future Work

Segmentation of fetal images is progressively gaining the interest of researchers around the world due to the edge-cutting image acquisition technology. Deep learning has been widely employed for medical image segmentation in the last few years. This review surveyed the latest deep learning techniques used for 2D medical image segmentation. The methodologies discussed in this review range from traditional analysis approaches to most modern Neural Network approaches. Each study is summarized, and the core principle of methods are presented in this manuscript. Deep learning has shown a state-of-the-art performance in the last five years, especially in ultrasound image segmentation. The existing research shows that almost 80% of deep learning models are based on Encoder-decoder networks and 2D fully convolutional neural networks. Though deep learning shows unprecedented performance enhancements in the field of fetal image segmentation, it has some limitations as well.

Most of the deep learning techniques involve a high cost in labeling the training dataset.

The heterogeneous appearance of the organ is also the biggest challenge in the segmentation of medical images with deep learning models.

Most of the deep learning models encounter the problem of overfitting due to the unavailability of large fetal imaging datasets.

Moreover, most AI researchers do not fully understand the specific clinical needs. To commercialize deep learning techniques, institutions should facilitate the collaboration between clinicians and AI scientists. Currently, supervised learning techhniues are widely employed in the segmentation of fetal 2D images and in other medical applications, as it is evident from the research presented. But very less research is being done on employing unsupervised ML into Medical image analysis. The future work should focus on incorporating unsupervised learning techniques for the segmentation of fetal images.

References

[1] S. J. Sree and C. J. C. M. I. Vasanthanayaki, "Ultrasound Fetal Image Segmentation Techniques: A Review," vol. 15, no. 1, pp. 52-60, 2019.

[2] N. Sharma and L. M. J. J. o. m. p. A. o. M. P. o. I. Aggarwal, "Automated medical image segmentation techniques," vol. 35, no. 1, p. 3, 2010.

[3] J. J. H. o. m. i. Rogowska, processing and analysis, "Overview and fundamentals of medical image segmentation," pp. 69-85, 2000.

[4] A. Norouzi et al., "Medical image segmentation methods, algorithms, and applications," vol. 31, no. 3, pp. 199-213, 2014.

[5] K. Irene, H. Haidi, N. Faza, and W. J. a. p. a. Chandra, "Fetal Head and Abdomen Measurement Using Convolutional Neural Network, Hough Transform, and Difference of Gaussian Revolved along Elliptical Path (Dogell) Algorithm," 2019.

[6] T. L. van den Heuvel, D. de Bruijn, C. L. de Korte, and B. v. J. P. o. Ginneken, "Automated measurement of fetal head circumference using 2D ultrasound images," vol. 13, no. 8, p. e0200412, 2018.

[7] R. Buettner, M. Bilo, N. Bay, and T. Zubac, "A Systematic Literature Review of Medical Image Analysis Using Deep Learning," in 2020 *IEEE Symposium on Industrial Electronics & Applications (ISIEA),* 2020, pp. 1-4: IEEE.

[8] S. M. Jardim, M. A. J. U. i. m. Figueiredo, and biology, "Segmentation of fetal ultrasound images," vol. 31, no. 2, pp. 243-250, 2005.

[9] S. Rueda et al., "Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge," vol. 33, no. 4, pp. 797-813, 2013.

[10] J. Perez-Gonzalez, J. B. Muńoz, M. R. Porras, F. Arámbula-Cosío, and V. Medina-Bańuelos, "Automatic fetal head measurements from ultrasound images using optimal ellipse detection and texture maps," in *VI Latin American Congress on Biomedical Engineering CLAIB* 2014, Paraná, Argentina 29, 30 & 31 October 2014, 2015, pp. 329-332: Springer.

[11] V. Shrimali, R. Anand, and V. Kumar, "Improved segmentation of ultrasound images for fetal biometry, using morphological operators," in 2009 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 459-462: IEEE.

[12] G. V. Ponomarev, M. S. Gelfand, and M. D. J. P. o. c. U. b. m. f. f. u. i. Kazanov, ISBI, "A multilevel thresholding combined with edge detection and shape-based recognition for segmentation of fetal ultrasound images," vol. 2012, pp. 17-19, 2012.

[13] W. Lu, J. Tan, R. J. U. i. m. Floyd, and biology, "Automated fetal head detection and measurement in ultrasound images by iterative randomized Hough transform," vol. 31, no. 7, pp. 929-936, 2005.

[14] J. Li et al., "Automatic fetal head circumference measurement in ultrasound using random forest and fast ellipse fitting," vol. 22, no. 1, pp. 215-223, 2017.

[15] Y. Meng et al., "CNN-GCN aggregation enabled boundary regression for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention,* 2020, pp. 352-362: Springer.

[16] L. Wu, Y. Xin, S. Li, T. Wang, P.-A. Heng, and D. Ni, "Cascaded fully convolutional networks for automatic prenatal ultrasound image segmentation," in 2017 *IEEE 14th international symposium on biomedical imaging (ISBI 2017)*, 2017, pp. 663-666: IEEE.

[17] B. M. Williams and Y. Zheng, "Improving fetal head contour detection by object localisation with deep learning," in *Medical Image Understanding and Analysis: 23rd Conference, MIUA* 2019, Liverpool, UK, July 24-26, 2019, Proceedings, 2020, p. 142.

[18] H. Ravishankar, S. M. Prabhu, V. Vaidya, and N. Singhal, "Hybrid approach for automatic segmentation of fetal abdomen from ultrasound images using deep learning," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI*), 2016, pp. 779-782: IEEE.

[19] Z. Sobhaninia et al., "Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning," in 2019 41st annual international conference of the IEEE engineering in *medicine and biology society (EMBC)*, 2019, pp. 6545-6548: IEEE.

[20] L. Wu, J.-Z. Cheng, S. Li, B. Lei, T. Wang, and D. J. I. t. o. c. Ni, "FUIQA: Fetal ultrasound image quality assessment with deep convolutional networks," vol. 47, no. 5, pp. 1336-1349, 2017.

[21] G. Wang et al., "Interactive medical image segmentation using deep learning with image-specific fine tuning," vol. 37, no. 7, pp. 1562-1573, 2018.

[22] S. S. M. Salehi et al., "Real-time automatic fetal brain extraction in fetal MRI by deep learning," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018),* 2018, pp. 720-724: IEEE.

[23] S. J. Sree and c. Vasanthanayaki, "Fetal ultrasound image segmentation using fuzzy connectedness algorithm.", *Advances in Mathematics: Scientific Journal* 9 (2020), no.10, 8293–8301.

[24] J. Li et al., "Automatic fetal brain extraction from 2D in utero fetal MRI slices using deep neural network," vol. 378, pp. 335-349, 2020.

[25] Y. Zeng, P.-H. Tsui, W. Wu, Z. Zhou, and S. J. J. o. D. I. Wu, "Fetal Ultrasound Image Segmentation for Automatic Head Circumference Biometry Using Deeply Supervised Attention-Gated V-Net," vol. 34, no. 1, pp. 134-148, 2021.

[26] S. Nurmaini *et al.*, "Deep learning-based computer-aided fetal echocardiography: application to heart standard view segmentation for congenital heart defects detection," *Sensors,* vol. 21, no. 23, p. 8007, 2021.

[27] R. Sreelakshmy *et al.*, "An Automated Deep Learning Model for the Cerebellum Segmentation from Fetal Brain Images," *BioMed Research International,* vol. 2022, 2022.