Fuzzy Attention U-Net Architecture Based Localization and YOLOv5 Detection for Fetal Cardiac Ultrasound Images

S. Satish[1*] and Dr. Herald Anatha Rufus[2]

[1*] Full Time Research Scholar, Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R& D Institute of Science and Technology, India.
satishsaiece@gmail.com, Orcid: https://orcid.org/0000-0002-4729-6676

[2] Associate Professor, Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R& D Institute of Science and Technology, India.
drrufus@veltech.edu.in, Orcid: https://orcid.org/0000-0003-0965-2796

Received: September 30, 2023; Accepted: December 25, 2023; Published: March 30, 2024

Abstract

The importance of early identification of congenital heart disease is highlighted by the fact that it accounts for around 28% of all congenital abnormalities this is the primary reason why fetus die. The necessity of having a thorough understanding of normal cardiac architecture has been highlighted by the quick advancements in fetal heart imaging techniques that have occurred in recent years. Without this information, it is challenging, even impossible, to distinguish the many manifestations of congenital heart illness. This research suggests an immediate fetal cardiac identification technique employing US pictures with the You Only Look Once v5 (YOLOv5) framework and localization using Fuzzy Attention U-Net (FAU-Net) framework in order to enhance the interpretation of the anatomy of the fetal heart through Ultrasound (US) for precise and instantaneous diagnoses. Localization is accomplished using a FU-Net architecture, which limits the training set to image-level plane labels. This is a crucial component of the proposed study since, for big datasets, it would take too much time to produce bounding box annotations, which are not always captured. The FAU-Net design has been optimized for best performance. The YOLOv5 framework is built to function in real-time and deliver the best results for object identification. With the aid of appropriate fine-tuning, it may function effectively to automatically identify tiny fetal cardiac objects in a fast phase. Depending on how much of a detection overlapped with ground-truth bounding boxes, it was determined if it was a genuine positive or a false positive. This research primarily aids medical professionals in the diagnosis of fetal cardiac anatomy. The efficiency of the outcomes is evaluated using metrics including accuracy, memory, average precision (AP), mean average precision (mAP), and F-measure.

Index Terms: Congenital Heart Disease (CHD), Ultrasound Imaging, Deep Learning, Object Detection, You Only Look Once v5 (YOLOv5), Fuzzy Attention U-Net (FAU-Net) Architecture, and Object Localization.

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 15, number: 1 (March), pp. 01-16. DOI: *10.58346/JOWUA.2024.I1.001*

*Corresponding author: Full Time Research Scholar, Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R& D Institute of Science and Technology, India.

1 Introduction

Congenital heart disease (CHD), which affects 2 to 6.5 out of every 1000 live births, is a significant contributor to newborn morbidity and death. The risk status of the mother has little effect on the prevalence of CHD (Nayak, K., 2016). The necessity for early medical as well as surgical intervention at or shortly after delivery in more than half of all CHD cases emphasizes the value of prenatal screening. Fetal echocardiogram is a specialized, in-depth ultrasound (US) test that assesses the heart's shape and function while the mother is still carrying the child in order to diagnose CHD and myocardial dysfunction before birth. When the embryonic heart has developed enough to be realistically visualized by transabdominal ultrasound, fetal echocardiogram is commonly done between 18 and 22 weeks of gestation. Transvaginal ultrasonography was once used to find the fetus heart development with a lower specificity from weeks 12 and 13, but the findings frequently needed to be confirmed by a thorough fetal echocardiogram.

Fetal echocardiography should be investigated pregnant women who have fetus or maternal conditions that might raise the chance of a heart abnormalities are present. In particular, echocardiography is an essential tool for identifying and treating CHD because it can accurately evaluate the embryonic heart's anatomy and function (Shi, C., 2002). The optimum image for prenatal diagnoses and examinations for fetal CHD, the developing baby's ultrasound Four-Chamber (FC) view, provides physicians with a clear picture of the fetal cardiac morphology (Pan, S., 2020). During the initial tests of fetal CHD, the key point of evaluation for the clinicians is the morphological and functional characteristics of the unborn heart (Pan, S., 2020). It is important to note that the division of tissues or lesions can statistically examine the clinical characteristics associated with volume or developmental morphology, aid physicians in correctly diagnosing the patient's condition, and plan an appropriate treatment approach (Pan, S., 2020). Prenatal CHD diagnosis has a substantial influence on the course of the pregnancy, the decision to abort, fetal treatment, delivery method, and the requirement for tertiary care.

It is extremely challenging to study these nine fetus cardiac substructures because of the small size of the heart of the fetus, the small size among the nine fetus heart substructures, the not fixed fetus placements, and categorization uncleared due to a resemblance of the chambers of the heart (Pinheiro, D.O., 2019) (Nurmaini, S., 2021). A Computer-Aided Diagnosis (CAD) approach that assists doctors in automatically locating embryonic cardiac objects has garnered a lot of interest in recent years (Nurmaini, S., 2021) (Madani, A., 2018) in an effort to alleviate these issues. These techniques can help doctors automatically identify fetal cardiac structures which can have a substantial impact on the early detection of congenital heart disorders. Furthermore, a method including automated fetal US interpretation can involve non-experts using portable equipment at point-of-care settings.

Using CAD and Artificial Intelligence (AI), fetal cardiac imaging analysis, it is feasible to autonomously segregate and categorize the developing cardiac organ (Torrents-Barrena, J., 2019) (Zhang, B., 2021) to find problems with the septum of the heart (Gandhi, S., 2018). Three distinct forms of holes in the atria, ventricles, or both are utilized to identify CHD: AVSD (atrioventricular septal defect), VSD (ventricular septal defect), ASD (atrial septal defect) (Nurmaini, S., 2020) (Mcleod, G., 2018). The situation is particularly dangerous because it allows blood to flow back and forth through the anterior chamber of the cardiovascular system into the left (Puri, K., 2017). Based on DL (Deep Learning), CNN (Convolutional neural network) architecture, is an AI technique that might be used to identify prenatal items (Torrents-Barrena, J., 2019) (Zhang, B., 2021).

Regarding CNN's capability in sorting, the process of segmentation and recognition using imaging for diagnosis several research have achieved impressive findings (Rezvy, S., 2020) (Vo, K., 2020). Applications like CNN acquire data and provide reliable forecasts along with judgements using previous information (Zhong, W., 2018). CNN performs adaption tasks without the use of specific programming. Regarding a fetal echocardiography study based on CNN (Torrents-Barrena, J., 2019), leaking via missing boundaries caused due to intra-chamber walls is still a problem. The You Only Look Once v5 (YOLOv5) framework used in this study is built to function in real-time while producing the best results for object recognition. To accurately diagnose the fetal heart, Fuzzy Attention U-Net (FAU-Net) has been established for object identification. Develop a system for automatically locating objects and detecting them in ultrasound pictures of both normal and pathological anatomical structures, such as ASD, VSD, and AVSD. It shows how automated localization along with detection techniques may greatly raise the rate from CHD diagnosis.

2 Literature Review

Nurmaini et al. (2021) offered the use of deep learning for computer-assisted fetus heart ultrasound testing utilizing an instance segmented technique, which naturally segment each of the four standard cardiac images while also identifying the abnormalities. Many tests are run using 1149 Predictions from fetus cardiac image 24 things, such as three occurrences of congenital heart defects, three typical morphologies of a heartbeat of a fetus, and 17 heart chamber elements in each perspective. The results demonstrated that the suggested model was successful in segmenting typical points of view, with a 79.97% overlap across collaboration and an 89.70% Dice coefficient similarity. It also did well in the detection of CHD, with mean average accuracy of about 98.30% for intra-patient variation and 82.42% for inter-patient variance. With the application of automated segmentation and detection techniques, the prevalence with congenital heart disease diagnosis might rise.

To enhance feature learning, Qiao et al. (2020) presented a Multistage Residual Hybrid Attention Module (MRHAM). Then, an enhanced YOLOv4 detection model for object detection called MRHAM-YOLOv4-Slim is proposed. In particular, the MRHAM-YOLOv4-Slim's backbone replaces the residual mapping of identity with the MRHAM, precisely finding the four crucial chambers in fetal FC images. Sapitri and Darmawahyuni (Sapitri, A.I., 2021) utilized a quicker Regional Convolutional Neural Network (R-CNN) using the R-CNN mask technique. devised for deep learning in instances (Bae, D., 2021). The aortic area of 151 ultrasound pictures of the heart of a fetus has been studied using the suggested method. Metrics on the object that was identified using a mean Average Precision (mAP) value of 83.71% were evaluated in order to test the evaluation findings.

You Only Look Once (YOLO) structure -based instantaneous embryonic heart structure identification using US video was proposed by Sapitri et al. (2023). The entire loop of neural network in YOLO concurrently predicts cardiac substructure objects, packages, and probabilities of classes. Forty fetal echocardiography recordings are prepared using the recently released YOLOv7 structure and then modified to function optimally and operate quickly in order to obtain dependable performance. The findings reached 17 frames per second (FPS) over nine cardiac structure entities in 0.3 ms, with the greatest mean average precision of 82.10%. Our study's key conclusion is that YOLOv7 can identify features of the embryonic cardiovascular structure in instantaneous even with a small sample size of US footage. With the aid of appropriate fine-tuning, such a network may function effectively to recognize tiny fetal cardiac objects autonomously in a quick phase. This research primarily aids medical professionals in the diagnosis of fetal cardiac anatomy.

An entirety Dilated Convolutional Chain W-Net module (DW-Net) was suggested by Xu et al. (2020) for precise dissection of seven significant anatomic features in the A4C view. The network consists of two parts: 1) a Dilated Convolutional Chain (DCC) for "gridding issue" minimization, multiscale contextual information aggregation, and precise localization of heart chambers. 2) A W-Net for better segmentation results and more accurate boundary determination. Extensive testing of the suggested technique on a dataset of 895 A4C views showed that DW-Net significantly outperformed certain well-known segmentation methods and could produce good segmentation outcomes, particularly Dice Similarity Coefficient (DSC), Pixel Accuracy (PA), and AUC.

Nurmaini et al.'s (2022) enhanced semantic segmentation method includes two processes—contour segmentation with U-Net framework and defect identification with Faster-RCNN architecture—and employs a specific recommendation network for septal defect detection. The model is trained using 764 ultrasound pictures from an apical four-chamber view, which contain three defective situations (namely, atrial septal defect, ventricular septal defect, and atrioventricular septal defect) and typical circumstances. It may correctly detect the heart of the fetus in both healthy and unhealthy conditions. Future prospects for the practical application of deep learning in the diagnosis of congenital cardiac conditions are quite promising.

Analysing whether computers can be taught to identify these viewpoints is the crucial first step Madani et al (2018) described as being necessary for thorough computer-assisted echocardiographic interpretation. CNN is used to categorize 15 standard views simultaneously (12 video, 3 still), based on annotated still photos and video over 267 transthoracic echocardiograms that recorded a variety of clinical variance in real-world settings. Reliability within 15 views was 91.70% even on a single low-resolution picture, compared to board-certified echocardiographers' range of 70.2-84.0%. The program classifies utilizing clinically pertinent picture attributes and can distinguish commonalities across related images, according to data visualization studies.

In freehand 2-D ultrasound images, 13 fetal standard views may be automatically detected by Baumgartner et al's unique CNN-based technique (Baumgartner, C.F., 2017), which can also localize the fetal anatomy using a bounding box. A significant addition is the network's ability to pinpoint the desired anatomy using only image-level labels and inadequate supervision. The architecture of the network is built to function in real-time and deliver the best results possible for the localization job. Results for frame recovery from stored movies in the past, real-time annotation, and localisation on a very big and difficult dataset made up of photos and videos taken during comprehensive clinical anomaly tests.

Shu et al. (2022) stated an Efficient Channel Attention U-Net (ECAU-Net) method for segmenting the cerebellum. The U-Net acts as the technique's segmentation framework, applying the encoding algorithm to acquire representations of features using utilizing the decoder to find segmentation results. One-dimensional layers of convolution with identical components are used in place of the complete connection layers in the traditional channel attention modules when combined with ECA modules. This significantly reduces the amount of model parameters without degrading performance. A significant addition is the network's ability to pinpoint the desired anatomy using only image-level labels and inadequate supervision. The structure of the network is built to function in real-time and deliver the best results possible for the localization job. Results for frame recovery from stored movies in the past, real-time annotation, and localization on a very big and difficult dataset made up of photos and videos taken during comprehensive clinical anomaly tests.

Shabanzadeh et al.'s (2022) proposal for a quick and accurate U-Net-based architecture for the job of segmenting medical images. Four adjusted 2D-convolutional, batch normalization, and 2D-transposed

convolutional layers make up the heart of the suggested U-Net model. A four-block encoder-decoder path is used. Fetal head circumference was measured using a publicly available dataset (HC18-Grand Challenge dataset), and the performance of our suggested architecture was evaluated using datasets developed specifically both head circumference as well as belly circumference segmentation tasks. swift also precise Just four necessary layers with adjusted parameters are employed for the process of encoding and decoding routes in the U-Net model, which is a finely tuned and well-structured version of the U-Net. The suggested design has a well-tuned structure that produces excellent segmentation accuracy while being quick.

3 Proposed Methodology

This paper provides an ultrasound-based real-time fetal cardiac detection algorithm using the You Only Look Once v5 (YOLOv5) framework. In YOLOv5, a neural network predicts cardiac substructure objects, boxes, and class probabilities simultaneously from start to finish. Using in order image-level plane labels known after training, the Fuzzy Attention U-Net (FAU-Net) Structure accomplishes localization. Results are assessed using metrics including reliability, recall, average precision (AP), mean average precision (mAP), and F-measure. The four fundamental processes are depicted in Figure 1 as object detection, localization, object detection, and result validation.

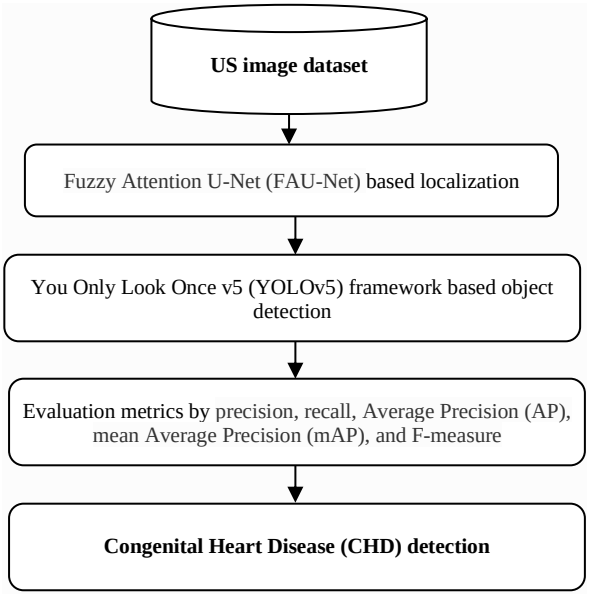

Figure 1: Diagram of the Proposed System's Flow

3.1. Image Acquisition and Fetal Cardiac Annotation

The first stage in doing studies regarding real-time fetal cardiac identification utilizing ultrasound is gathering a well-defined dataset. In this study, 540 normal, VSD, and AVSD US pictures of patients receiving regular pregnancies testing during the second trimester of pregnancy within the obstetrics and genecology section were collected. With consent waived for compliance, two knowledgeable experts identified all US photos. The decreased size of the fetal heart or the loss of distinguishing traits may be to blame for the model's difficulty in detecting fetal cardiac substructure objects (Madani, A., 2018). The medical record on every US film was examined to identify the fetus heart normal anatomy, and the specialists commented the chosen fetal US for educational reasons. The initial Digital Imaging and

Communication in Medicine (DICOM) standard was used to store all pictures in the US. The fetal cardiac substructures annotated procedure is carried out item by object. After the annotating process, Figure 2 dubbed all the data the ground truth box.

Normal VSD AVSD

Figure 2: Ground Truth Box of Fetal Cardiac Image Types

3.2. Fuzzy Attention U-Net (FAU-Net) based localization

Present the FAU-Net deep learning network, which is inspired by the U-Net network topology as well as attention mechanism. Figure 3 displays the structure in its entirety. Define "Block(x)", a function that twice does a 33 convolution, a batch normalization, and a ReLU activation. The output channel number, x, is referred to. The encoder part's job is to obtain multi-level compressed expressions of the image features by extracting features from the US picture. The 2 2 max-pooling procedure is used for down-sampling. A smaller US picture size and twice as many feature channels are added at each downsampling. The decoder part's function is to gradually restore the spatial dimensions along with details of the fetal cardiac US picture in accordance with the image characteristics, and to get output of image mask. Using bilinear interpolation, upsampling is accomplished. The prediction of each pixel's class is then applied using a 1 1 convolutional layer, designated as Conv(1 1, C), wherein C indicates the number of classes. C is set to 2 for localization of the picture. The encoder and decoder components are structurally symmetrical. The related upsampling and downsampling feature maps are linked by the copy operation. High-level and low-level features are combined in the feature map, as well as multi-level features are fused.

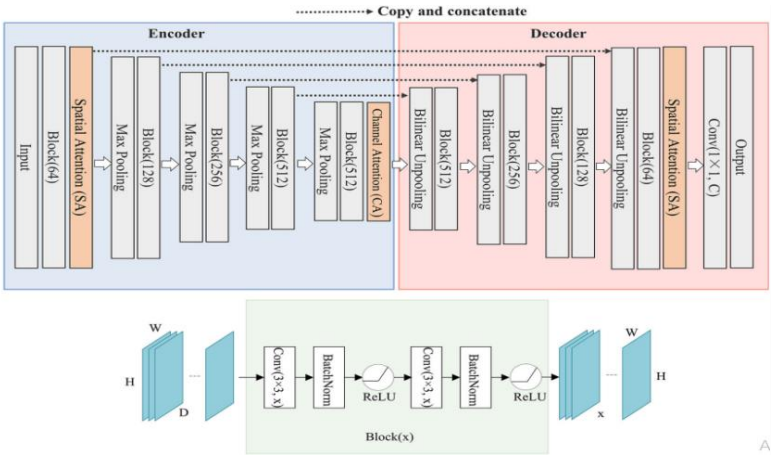

Figure 3: Structure of FAU-Net Framework

Four components make up the overall structure: an encoder, a decoder, a spatial attention module, and a channel attention module. The size of the output of Block(x), assuming an input map of features of size DHW, is DHW, where W indicates the feature map's width, H its height, D its input channel number, and x its output channel number. Compared to other pictures, the structure of the fetal cardiac image is more straightforward and stable in nature. The shot attitude and positioning are constant for gland slices, and the glands with a comparable degree of differentiation frequently have a similar form. The network makes use of the simple to use attention to space module and the attention channel module. They are influenced through the study of Squeeze-and-excitation (SE) (Hu, J., 2018) and Convolutional Block Attention Module (CBAM) (Sha, J., 2023). The backdrop will not be noticed since attention will be drawn to the items. For the segmentation job, the fuzzy edge was the most important, therefore the model will focus more on the margins of the glands.

Spatial Attention: The channel information is disregarded by spatial attention, which evaluates all channel properties identically. Since low-level feature maps primarily extract spatial features like contours and edges using fewer channels, spatial attention modules are employed with these maps. Figure 4 depicts the organizational layout within the spatial attention module. First, provide the aggregation operation the feature map $U \in \mathbb{R}^{C\times H\times W}$, which creates a spatial descriptor $p \in \mathbb{R}^{H\times W}$ with aggregate the feature map within its channel dimension (C).

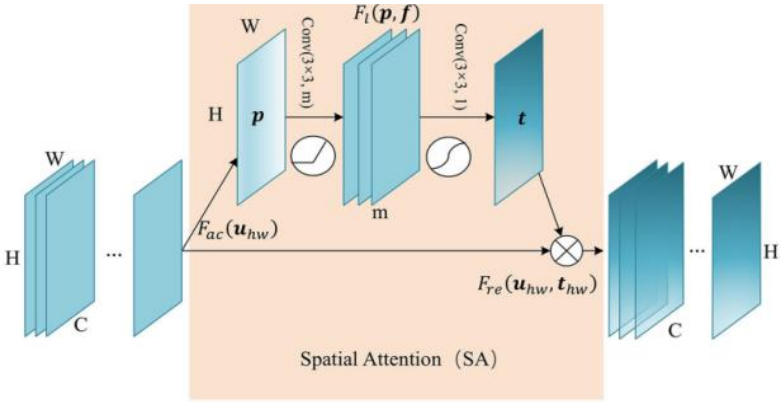

Figure 4: Spatial Attention (SA) Module Structure

A spatial descriptor called $p \in \mathbb{R}^{H\times W}$. is produced by the aggregate function F_{ac}. The spatial weights map $t \in \mathbb{R}^{H\times W}$ is produced via two convolutional layers using the self-learning function Fl. Finally, t is used by function F_{re} to produce the SA module's output. It produces a distribution of spatial attributes that is global,

$$p_{hw} = F_{ac}(u_{hw}) = \frac{1}{C}\sum_{i=1}^{C} u_{hw}(i) * f_{we} \quad (1)$$

where f_{we} is referred to represent the fuzzy weight for level and $u_{hw} \in \mathbb{R}^{C}$ refers towards the particular feature at spatial point (h,w). The value of a location in the regional window is x_{hw}, while its associated trigonometric function membership is μ_{hw}, as given in equation (2). This is done by computing the function of membership of every pixels as the weight of weighted entropy.

$$\mu_{hw} = \frac{\sin\left[\frac{\pi\left[1-\frac{x_{max}-x_{hw}}{K}\right]}{2}\right]+1}{2} \quad (2)$$

Equation (2), where $0.5(x_{max} - x_{min}) < K < x_{max}$. x_{max} denotes the range's largest value and x_{min} denotes the range's minimum value. Each feature point's entropy weight, ew_{hw} is determined as given in equation (3).

$$ew_{hw} = \frac{\mu_{hw}}{\sum_{k=1}^{n} \mu_{hw}} \quad (3)$$

where each fuzzy subset's feature number, n, is given. If the area window is $h \times w$ in size, then n= $h \times w$. Finally, as stated in Equation (4), the fuzzy weighted entropy H associated with every fuzzy subset is determined.

$$H = \sum_{k=1}^{n} [ew_{hw} e^{1-ew_{hw}} + (1 - ew_{hw}) e^{ew_{hw}}] \quad (4)$$

For channel dimensions, the aggregate function $\boldsymbol{F_{ac}}$ provides global average pooling. A weight self-education technique comes next. Layers of convolution are used to implement it. The spatial weights map $t \in \mathbb{R}^{H \times W}$ is generated adapted by the function $F_l\ (p, f)$, which seeks to completely represent the spatial correlation. The following is the estimating formula:

$$t = F_l(p, f) = \sigma\big(g(p, f)\big) = \sigma\big(f_2 \delta(f_1, p)\big) \quad (5)$$

where f_1 is the 3 × 3 convolution specified by Conv(3×3, m) and f_2 is the 3× 3 convolution described by Conv(3 × 3, 1). The hidden feature map's channel number is m. The activation function in question is ReLU, and the spatial weight created at position (h,w) using a sigmoid activation function is $t_{hw} \in (0, 1)$. Convolution is simply a spatial-wise self-attention function able to capture non-linear spatial interactions. It takes the original spatial description as input. The weights chosen in the previous phase are applied to feature map U. During spatially wise recalibration $F_{re}(u_{hw}, t_{hw})$, the feature values of various places in U are multiplied by various weights to create the output U ′ of the SA module.

$$u'_{hw} = F_{re}(u_{hw}, t_{hw}) = u_{hw} \cdot t_{hw} \quad (6)$$

Channel Attention: The channel attention module appears as a final layer of the encoder while the high-level function map primarily communicates complex features with a wide receptive field and extra channels. Through learning to utilize global information, this process enables the network to undertake feature recalibration, selecting enhancing valuable characteristics and limiting worthless features. Figure 5 depicts the channel attention module's organizational structure.

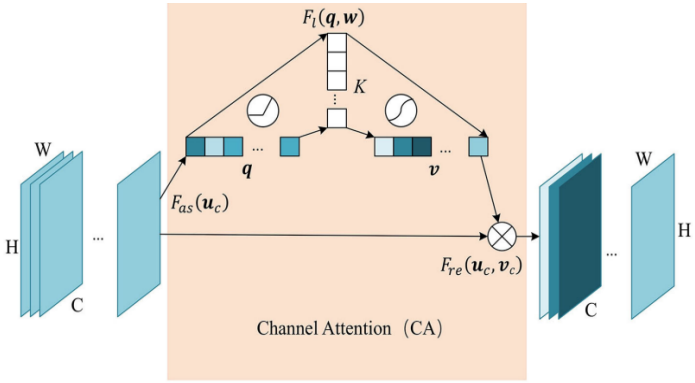

Figure 5: Channel Attention (CA) Module Layout

The aggregate function F_{as} generates the channel descriptor $q \in \mathbb{R}^C$. The self-learning function Fl, which is executed by two totally linked layers, generates the channel weights map $v \in \mathbb{R}^C$. Function Fre makes use of v to generate the CA module's outcome. Initially provide the aggregation operation the

feature map $U \in \mathbb{R}^{C\times H\times W}$, which aggregates the feature map in its spatial dimension (H × W) to produce a channel descriptor $q \in \mathbb{R}^{C}$. It results in a global dispersion of channel properties,

$$q_c = F_{as}(u_c) = \frac{1}{\mathrm{H} \times \mathrm{W}} \sum_{i=1}^{H} \sum_{i=j}^{W} u_c(i,j) \quad (7)$$

where the notation $u_c \in \mathbb{R}^{H\times W}$ designates the local channel c feature. The aggregate function F_{as} uses global average pooling for the spatial dimension. A weight self-learning mechanism follows. To implement it, fully connected layers are utilized. The $F_l(q,w)$ function, that creates the channel weights map $v \in \mathbb{R}^{C}$ in an adaptive manner with the goal of completely capturing the interactions between channels. The following is the calculating procedure:

$$v = F_l(q,w) = \sigma\big(g(q,w)\big) = \sigma\big(w_2\delta(w_1 q)\big) \quad (8)$$

where $w_1 \in \mathbb{R}^{K\times C}, w_2 \in \mathbb{R}^{C\times K}$ Number of hidden neurons is denoted by K. For channel c, the channel weights $v_c \in (0,1)$ are produced using a σ sigmoid activation function. It can record how non-linearly channels interact when their hidden layers are completely coupled. On the feature map U, the weight determined in the preceding step is applied. The output U′ for the CA module is created by multiplying features values of different channels in U by varying weights by channel-wise recalibration$F_{re}(u_c, v_c)$,

$$u_c' = F_{re}(u_c, v_c) = u_c . v_c \quad (9)$$

For picture localisation in CHD images, several procedures have been utilized.

3.3. YOLOv5 based Object Detection

Three components make up the bulk of YOLOv5: the head, neck, and backbone. Figure 6 depicts the YOLOv5s network's design. The backbone is in charge of removing feature data from photos. In order to reduce the number of network layers and factors whereas expanding the base-level receptive field while preserving the extraction of feature accuracy to the greatest extent possible, the network's first layer uses a convolution module with a 6 × 6 massive convolutions kernel that converts the dimension and height information related to the image into channel information. In this network, the C3 module, that effectively functions as a residual module, is crucial for extracting features. The input data is split into two halves; one section undergoes the bottle neck module for obtaining deep features, and the other section only goes through one convolution module. The feature extraction is finally completed by fusing the two components. By using multistage max-pooling to gather local features at various sizes, the Spatial Pyramid Pooling - Fast (SPPF) module concatenates them to expand the receptive field while maintaining the size of the feature map. The collar part employs the Path Aggregation Network (PANet) framework, which combines low-level localization features and high-level semantic features acquired by the main network via top-down and bottom-up pathways to enhance the detection of objects of various sizes and integrate them into heads of various scales. Information output is handled by the head. To correlate with various item sizes, different scale heads have distinct feature map sizes. Each grid in the feature map generates a preset piece of data, such as the projected category, item confidence, bounding box center coordinates, width, and height. These picked cells are then filtered using the non-maximum suppression (NMS) approach according to their confidence and location data, producing the predicted items.

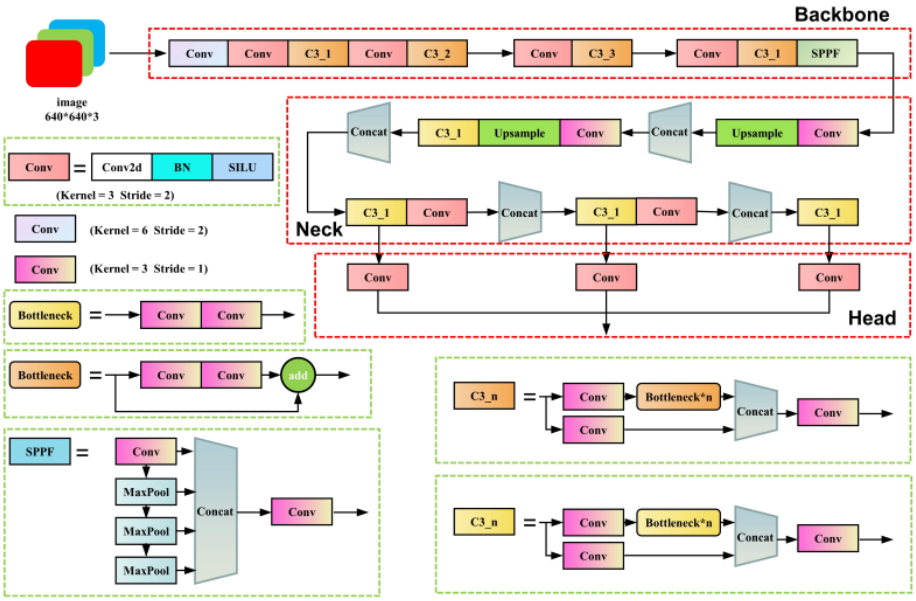

Figure 6: YOLOv5 Network Structure

Figure 6 displays the detailed design for each component as green boxes and the network topology as red boxes. The box loss, object loss, and class loss are the three components that make up the loss function in YOLOv5. Binary cross-entropy reduction is used to determine the object's reduction and class reduction, whereas complete IoU (CIoU) loss is used to estimate the box loss. YOLOv5 is separated as various models, that range from the model with the fewest features through the model with the most: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, as well as YOLOv5x, depending on the depth and breadth of the network. YOLOv5s is chosen as the basis model later relating the responsiveness and correctness of several models.

4 Results and Discussion

The real-time recognition of fetus cardiovascular structure is examined in this section using the YOLOv5 framework. Each network was implemented, tested, and trained using the MATLABR2021a software, and the following specifications were followed: These requirements apply to Ubuntu 18.04: Intel(R) Xeon(R) Platinum 8255C CPU, 43 GB RAM, RTX 2080 Ti GPU. On the aforementioned platform, both the models utilized in this research and the models utilized as an example have been trained and evaluated. The information collection was utilized by the framework that was recently developed through interpretation in order to more accurately assess the effectiveness of the model. Although the featured point positions inside the photos were partially obscured by extremely noisy welding, the algorithm continues to detect the locations of all points of feature by evaluating the global features. In noisy environments, the framework can maintain high accuracy and robustness because to this. Frames per second (FPS) have also been used to gauge the model's performance. Precision, recall, AP, mAP, and F-measure measures have each been used to assess the framework's real-time abilities and accuracy in detecting (Sha, J., 2023) (Du, Z., 2019). Although recall is used to confirm the existence of each item, precision examines how well the object estimates its location. They use the following calculus formulas:

$$Precision = \frac{TP}{TP + FP} \qquad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$F - measure = \frac{2.precision.recall}{precision + recall} \quad (12)$$

$$mAP = \frac{1}{N}\sum_{c=1}^{N} AP_c \quad (13)$$

In addition to Equation (10-13) as well as Equation 10-13, samples are considered to be accurately predicted if they have annotation boxes that are close to the expected boxes and with Intersection over Union (IoU) value larger than the predetermined IoU threshold. The specimens that must be and are correctly classified as positive are denoted by the initials TP, FP, and FN, respectively. Specimen that should deserve to be categorized as negative but were instead classified as positive are denoted by the letters FP, whereas specimen that the ought to be classified as favorable but were instead classified as negative are denoted by the characters FP, FN, and TP. Equation (13), which gives the AP value for class c, displays a number of categories N as well as the area AP within the precision-recall (P-R) curve. The IoU threshold ranges from 0.5 to 0.95, and the mAP 0.5:0.95 value indicates the average mAP value during this range. With the IoU threshold is set to 0.5, the mAP value is that value. The mAP assesses the model's recognition of N categories. Item detection accuracy and recall are totally and correctly reflected by the mAP. Therefore, within the attributes of the point identification task examined in this work, the smaller the missed as well as incorrect rate of detection using the visuals, the larger the mAP, which largely reflects improved detection accuracy. YOLOv4Slim, MRHAM-YOLOv4Slim, Supervised Object detection using Normal data Only (SONO)-YOLOv2 (Komatsu, M., 2021), and YOLOv5 are examples of convolutional block attention modules.

Table 1: Performance Comparison with the Methods

Methods	Precision (%)	Recall (%)	F-measure (%)	AP (%)	mAP(%)	FPS
CBAM- YOLOv4Slim	79.25	80.71	78.39	78.19	80.22	75
MRHAM-YOLOv4Slim	84.87	83.66	81.55	82.55	83.84	51
SONO-YOLOv2	88.91	87.69	85.57	86.87	87.61	42
YOLOv5	93.33	94.33	93.35	92.40	92.41	28

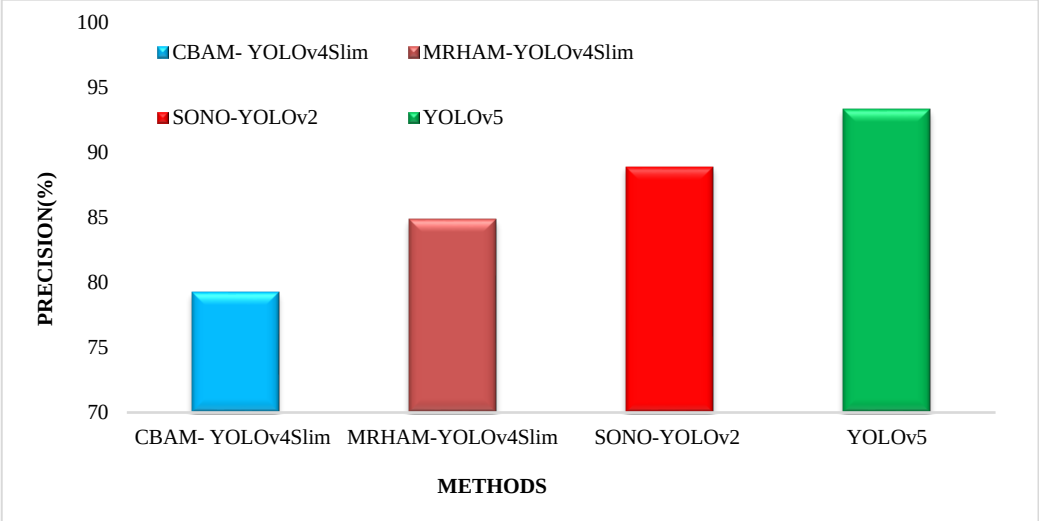

Figure 7: Precision Results Comparison Vs. Object Detection Methods

The precision outcomes comparing among multiple objects detecting techniques is shown in Figure 7. The precision of object identification techniques such as CBAM-YOLOv4Slim, MRHAM-YOLOv4Slim, SONO-YOLOv2, and YOLOv5 has been evaluated. Results for accuracy are 79.25%, 84.87%, 88.91%, and 93.33% for approaches like CBAM-YOLOv4Slim, MRHAM-YOLOv4Slim, SONO-YOLOv2, and YOLOv5.

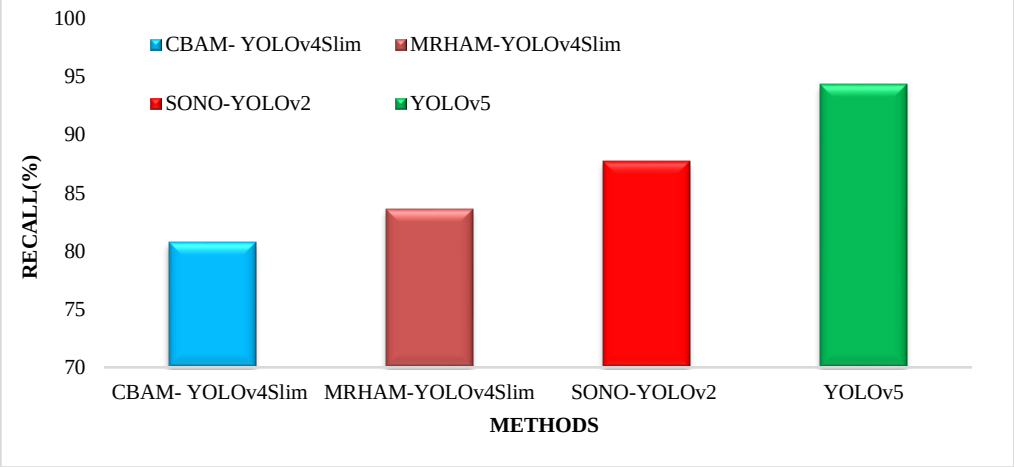

Figure 8: Recall Results Comparison Vs. Object Detection Methods

Figure 8 compares the recall outcomes of several object detecting techniques. In the CHD dataset, the recall of object identification techniques such CBAM- YOLOv4Slim, MRHAM-YOLOv4Slim, SONO-YOLOv2, and YOLOv5 has been evaluated. Recall rates of 80.71%, 83.66%, 87.69%, and 94.33% are obtained using the techniques CBAM-YOLOv4Slim, MRHAM-YOLOv4Slim, SONO-YOLOv2, and YOLOv5.

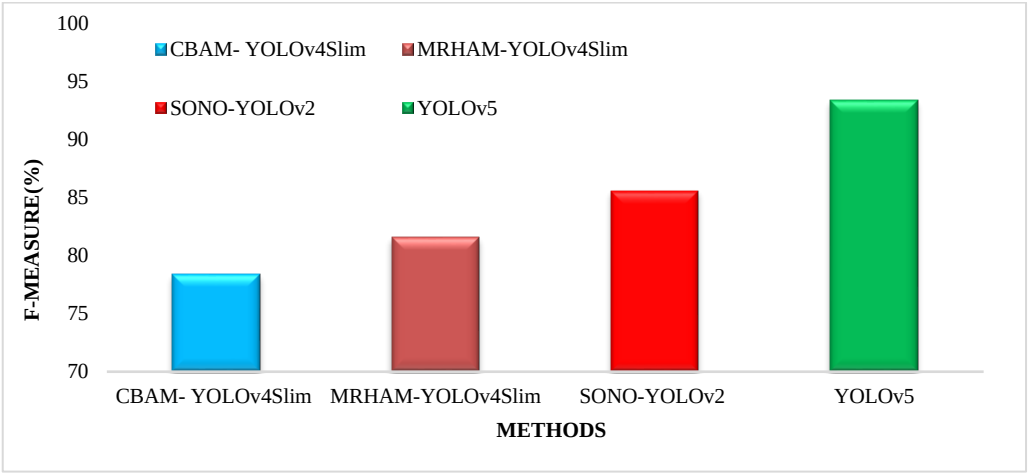

Figure 9: F-measure Results Comparison Vs. Object Detection Methods

Figure 9 shows a comparison of detecting objects algorithms using F-measure findings. The performance of object identification techniques like CBAM-YOLOv4Slim, MRHAM-YOLOv4Slim, SONO-YOLOv2, and YOLOv5 has been evaluated using the CHD dataset's f-measure. The f-measure findings for the techniques CBAM-YOLOv4Slim, MRHAM-YOLOv4Slim, SONO-YOLOv2, and YOLOv5 are 78.39%, 81.55%, 85.57%, and 93.35%, respectively.

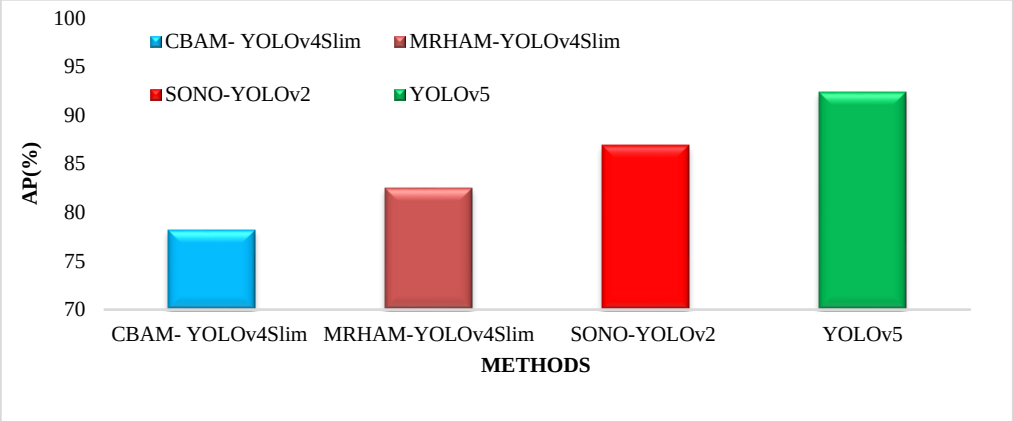

Figure 10: AP Results Comparison Vs. Object Detection Methods

Figure 10 shows an analysis of object detecting techniques' AP findings. In the CHD dataset, object identification techniques including CBAM-YOLOv4Slim, MRHAM-YOLOv4Slim, SONO-YOLOv2, and YOLOv5 have been evaluated by AP. CBAM-YOLOv4Slim, MRHAM-YOLOv4Slim, SONO-YOLOv2, and YOLOv5 techniques yield AP values of 78.19%, 82.55%, 86.87%, and 92.40%, respectively.

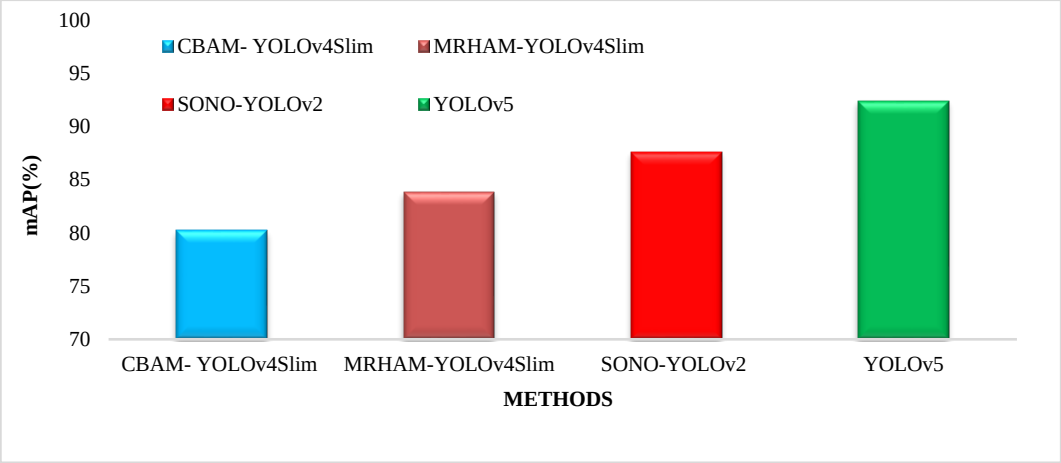

Figure 11: mAP Results Comparison Vs. Object Detection Methods

Figure 11 shows a contrast of detection of objects techniques using mAP data. The mAPin CHD dataset has evaluated object identification techniques including CBAM- YOLOv4Slim, MRHAM-YOLOv4Slim, SONO-YOLOv2, and YOLOv5. The mAP values for the techniques CBAM-YOLOv4Slim, MRHAM- YOLOv4Slim, SONO- YOLOv2, and YOLOv5 are 80.22%, 83.84%, 87.61%, and 92.41%, respectively

5 Conclusion and Future Work

This study uses the You Only Look Once v5 (YOLOv5) framework, which is built to run in real-time and produce the best results for object recognition. To accurately diagnose the fetal heart, Fuzzy Attention U-Net (FAU-Net) was recently established for object localisation. Create a system for automated localization and object recognition in both normal and aberrant anatomical structures including ASD, VSD, and AVSD using ultrasound pictures. The Spatial Pyramid Pooling - Fast (SPPF)

aspect in the YOLOv5 framework gathers local features of different sizes and combines these using multistage max-pooling to expand the field of reception while altering the extent of each feature map. By combining the low-level localization data obtained from the backbone network with the high-level semantic features, the neck segment implements a Path Aggregation Network (PANet) architecture via top-down and bottom-up pathways. Even with a small training sample of US images, the YOLOv5 model demonstrated favourable outcomes for fetal cardiac identification, indicating its potential for fetal cardiac recognition. The validation outcomes show that the model may be used to a variety of fetal heart anatomies. The suggested model has good promise for CHD with three defective components and promising fetal cardiac detection. The automated categorization of fetal cardiac models will be created in the future. When a fetal heart is examined during pregnancy, the categorized system will automatically identify if it has a septal defect based on the fetal cardiac region that is important that YOLOv5 has retrieved.

References

[1] Bae, D., & Ha, J. (2021). Performance Metric for Differential Deep Learning Analysis. *Journal of Internet Services and Information Security (JISIS)*, *11*(2), 22-33.

[2] Baumgartner, C.F., Kamnitsas, K., Matthew, J., Fletcher, T.P., Smith, S., Koch, L.M., & Rueckert, D. (2017). SonoNet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. *IEEE transactions on medical imaging*, *36*(11), 2204-2215.

[3] Du, Z., Yin, J., & Yang, J. (2019). Expanding receptive field yolo for small object detection. In *Journal of Physics: Conference Series*, *1314*(1), 1-6. IOP Publishing.

[4] Gandhi, S., Mosleh, W., Shen, J., & Chow, C.M. (2018). Automation, machine learning, and artificial intelligence in echocardiography: a brave new world. *Echocardiography*, *35*(9), 1402-1418.

[5] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132-7141.

[6] Komatsu, M., Sakai, A., Komatsu, R., Matsuoka, R., Yasutomi, S., Shozu, K., & Hamamoto, R. (2021). Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning. *Applied Sciences*, *11*(1), 1-12.

[7] Madani, A., Arnaout, R., Mofrad, M., & Arnaout, R. (2018). Fast and accurate view classification of echocardiograms using deep learning. *NPJ digital medicine*, *1*(1), 1-8.

[8] Madani, A., Arnaout, R., Mofrad, M., & Arnaout, R. (2018). Fast and accurate view classification of echocardiograms using deep learning. *NPJ digital medicine*, *1*(1), 1-8.

[9] Madani, A., Arnaout, R., Mofrad, M., & Arnaout, R. (2018). Fast and accurate view classification of echocardiograms using deep learning. *NPJ digital medicine*, *1*(1), 1-8.

[10] Mcleod, G., Shum, K., Gupta, T., Chakravorty, S., Kachur, S., Bienvenu, L., & Shah, S.B. (2018). Echocardiography in congenital heart disease. *Progress in cardiovascular diseases*, *61*(5-6), 468-475.

[11] Nayak, K., GS, N. C., Shetty, R., & Narayan, P. K. (2016). Evaluation of fetal echocardiography as a routine antenatal screening tool for detection of congenital heart disease. *Cardiovascular diagnosis and therapy*, *6*(1), 44-49.

[12] Nurmaini, S., Rachmatullah, M.N., Sapitri, A.I., Darmawahyuni, A., Tutuko, B., Firdaus, F., & Bernolian, N. (2021). Deep learning-based computer-aided fetal echocardiography: application to heart standard view segmentation for congenital heart defects detection. *Sensors*, *21*(23), 1-20.

[13] Nurmaini, S., Rachmatullah, M.N., Sapitri, A.I., Darmawahyuni, A., Jovandy, A., Firdaus, F., & Passarella, R. (2020). Accurate detection of septal defects with fetal ultrasonography images using deep learning-based multiclass instance segmentation. *IEEE Access*, *8*, 196160-196174.

[14] Nurmaini, S., Tama, B.A., Rachmatullah, M.N., Darmawahyuni, A., Sapitri, A.I., Firdaus, F., & Tutuko, B. (2022). An improved semantic segmentation with region proposal network for cardiac defect interpretation. *Neural Computing and Applications*, *34*(16), 13937-13950.

[15] Pan, S., & Luo, G. (2020). Application prospect of medical artificial intelligence in fetal echocardiography. *Chinese Journal of Practical Pediatrics*, *35*(11), 850-853.

[16] Pinheiro, D.O., Varisco, B.B., Silva, M.B.D., Duarte, R.S., Deliberali, G.D., Maia, C.R., & Beitune, P.E. (2019). Accuracy of prenatal diagnosis of congenital cardiac malformations. *Revista Brasileira de Ginecologia e Obstetrícia*, *41*, 11-16.

[17] Puri, K., Allen, H.D., & Qureshi, A.M. (2017). Congenital Heart Disease. *Pediatr. Rev*, *38*, 471-486.

[18] Qiao, S., Pang, S., Luo, G., Pan, S., Wang, X., Wang, M., & Chen, T. (2020). Automatic detection of cardiac chambers using an attention-based YOLOv4 framework from four-chamber view of fetal echocardiography, 1-9.

[19] Rezvy, S., Zebin, T., Braden, B., Pang, W., Taylor, S., & Gao, X.W. (2020). Transfer learning for Endoscopy disease detection and segmentation with mask-RCNN benchmark architecture. *In CEUR Workshop Proceedings*, *2595*, 68-72. CEUR-WS.

[20] Sapitri, A.I., & Darmawahyuni, A. (2021). Aorta Detection with Fetal Echocardiography Images Using Faster Regional Convolutional Neural Network (R-CNNs). *Computer Engineering and Applications Journal*, *10*(2), 115-124.

[21] Sapitri, A.I., Nurmaini, S., Rachmatullah, M.N., Tutuko, B., Darmawahyuni, A., Firdaus, F., & Islami, A. (2023). Deep learning-based real time detection for cardiac objects with fetal ultrasound video. *Informatics in Medicine Unlocked*, *36*.

[22] Sha, J., Wang, J., Hu, H., Ye, Y., & Xu, G. (2023). Development of an Accurate and Automated Quality Inspection System for Solder Joints on Aviation Plugs Using Fine-Tuned YOLOv5 Models. *Applied Sciences*, *13*(9), 1-19.

[23] Shabanzadeh, A., Sirjani, N., Akhavan, A., Shiri, I., Arabi, H., & Tarzamni, M. K. (2022). Fast and Accurate U-Net Model for Fetal Ultrasound Image Segmentation. *Ultrasonic Imaging*, 25-38.

[24] Shi, C., Song, L., Li, Y., & Dai, S. (2002). Value of four-chamber view of the fetal echocardiography for the prenatal diagnosis of congenital heart dise. *Zhonghua fu chan ke za zhi*, *37*(7), 385-387.

[25] Shu, X., Chang, F., Zhang, X., Shao, C., & Yang, X. (2022). ECAU-Net: Efficient channel attention U-Net for fetal ultrasound cerebellum segmentation. *Biomedical Signal Processing and Control*, *75*.

[26] 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. *Medical Image Analysis*, *51*, 61-88.

[27] 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. *Medical Image Analysis*, *51*, 61-88.

[28] Vo, K., Le, T., Rahmani, A. M., Dutt, N., & Cao, H. (2020). An efficient and robust deep learning method with 1-D octave convolution to extract fetal electro cardiogram. *Sensors*, *20*(13), 1-13.

[29] Xu, L., Liu, M., Shen, Z., Wang, H., Liu, X., Wang, X., & He, Y. (2020). DW-Net: A cascaded convolutional neural network for apical four-chamber view segmentation in fetal echocardiography. *Computerized Medical Imaging and Graphics*, *80*.

[30] 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), 1-15.

[31] Zhong, W., Liao, L., Guo, X., & Wang, G. (2018). A deep learning approach for fetal QRS complex detection. *Physiological measurement*, *39*(4), 1-10.

Authors Biography

S. Satish, currently working as full-time research scholar in Electronics and Communication Engineering Department at Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Chennai, Tamil Nadu, India. He received his Bachelor of Technology degree in Electronics and Communication Engineering in 2010 at Sri Manakula Vinayagar Engineering College under Pondicherry University, Puducherry and his Master of Technology degree in Wireless Communication at Pondicherry Engineering College, Puducherry. His area of interest is image processing, wireless communication and signal processing.

Dr.N. Herald Anantha Rufus is working as an Associate professor in Electronics and Communication Engineering Department at Vel Tech Rangarajan Dr.Sagunthala R & D Institute of Science and Technology, Chennai, Tamil Nadu, India. He received his BE degree in Electronics and Communication Engineering in 2000 at Noorul Islam College of Engineering under Manonmaniam Sundaranar University, Tirunelveli and his ME degree in Digital Communication and Networking at Arulmigu Kalasalingam College of Engineering under Madurai Kamaraj University, Madurai. He obtained his doctoral degree in the specialization of information and communication from Anna University, Chennai. He has teaching experience of 19 years and guided many UG and PG projects. He has authored various books on “Design of Electrical Machines” and “Artificial Intelligence for Beginners”. His area of interest is image processing and soft computing techniques, and he has published many research papers in international journals and national conferences.