Ingénierie des Systèmes d'Information
Vol. 30, No. 5, May, 2025, pp. 1363-1372
Journal homepage: http://iieta.org/journals/isi

Enhancing Medical Image Instance Segmentation Using Histogram Equalization and Blind Deblurring: A Preliminary Study

Hadi Syaputra[1,2], Siti Nurmaini[3*], Radiyati Umi Partan[4], Muhammad Taufik Roseno[1,2]

[1] Doctoral Program in Engineering Faculty of Engineering, Universitas Sriwijaya, Indralaya 30662, Indonesia
[2] Faculty of Computer Science, Universitas Sumatera Selatan, Palembang 30128, Indonesia
[3] Intelligent System Research Group, Universitas Sriwijaya, Palembang 30128, Indonesia
[4] Faculty of Medicine, Universitas Sriwijaya, Indralaya 30662, Indonesia

Corresponding Author Email: siti_nurmaini@unsri.ac.id

https://doi.org/10.18280/isi.300521

Received: 25 April 2025
Revised: 19 May 2025
Accepted: 27 May 2025
Available online: 31 May 2025

Keywords:
instance segmentation, Mask R-CNN, fetal heart, ultrasound imaging, HE, AHE, CLAHE, blind deblurring, IoU, DSC, mAP

ABSTRACT

Accurate instance segmentation of fetal heart structures in ultrasound (USG) images remains challenging due to low image quality and the small size of cardiac components. This study evaluates the impact of various image enhancement techniques on segmentation performance using the Mask R-CNN framework. We investigated Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), Contrast Limited Adaptive Histogram Equalization (CLAHE), Blind Deblurring (BD), and their combinations as preprocessing steps. A total of 176 clinically annotated ultrasound images were used, encompassing ten anatomical classes of the fetal heart. Twenty-two models were trained using different enhancement strategies and momentum values. Among them, Model 14 using combined AHE and CLAHE with a momentum of 0.9 achieved the best performance, with a mean Average Precision (mAP) of 0.3049 ± 0.0184, Intersection over Union (IoU) of 0.5887 ± 0.0366, and Dice Similarity Coefficient (DCS) of 0.7032 ± 0.0382. These findings highlight the effectiveness of local contrast enhancement in segmenting small and complex anatomical regions. Integrating complementary enhancement techniques can significantly improve segmentation quality and support more accurate fetal cardiac assessment in clinical settings.

1. INTRODUCTION

Medical Image Instance Segmentation (MIIS) represents an advanced stage in medical image processing that aims to classify every pixel in an image and distinguish between individual objects belonging to the same class. This process produces precisely segmented outputs, each representing an anatomical structure, organ, or pathological region within the medical image [1, 2]. The segmentation results provide critical spatial information such as boundaries, locations, sizes, and counts of objects, serving as valuable references for medical diagnosis and clinical decision-making [3, 4].

However, MIIS remains technically challenging, with several factors negatively affecting segmentation accuracy. Key issues include the generally low quality of medical images, limited availability of annotated datasets, the diverse visual patterns of bodily structures in medical imaging, and various difficulties in image preprocessing stages [5-8]. Therefore, more innovative and flexible approaches are essential to produce more accurate and reliable segmentation outcomes.

In recent decades, various techniques have been investigated for MIIS, ranging from classical image processing methods such as edge detection, thresholding, region growing, and clustering to more recent machine learning (ML) and deep learning (DL) approaches [9]. Among DL-based methods, Mask R-CNN has emerged as one of the most effective frameworks for instance segmentation. As an extension of Faster R-CNN, Mask R-CNN introduces an additional branch to generate pixel-level masks for each Region of Interest (RoI), thereby enabling precise detection and delineation of individual anatomical instances, even in cases involving overlap or indistinct boundaries [10].

Building upon the success of Convolutional Neural Network (CNN)-based methods, various models have been proposed to enhance the performance of MIIS [11]. A key advantage of CNNs is their capability to extract complex features directly from raw medical images. These models can be trained under both supervised and unsupervised learning paradigms [12]. In this study, a supervised learning approach is employed, wherein the Mask R-CNN model is trained on annotated datasets to facilitate accurate identification of anatomical structures and their spatial characteristics.

Despite recent advancements, a key limitation of CNN-based models such as Mask R-CNN is their sensitivity to image quality. These models often underperform when applied to medical images characterized by low contrast, noise, or blurring common artifacts in modalities such as ultrasound and X-ray imaging [13-15]. While several studies have sought to address these challenges, they often lack a comprehensive

preprocessing pipeline or fail to systematically evaluate the impact of multiple image enhancement techniques on instance segmentation performance.

To address this gap, this study integrates a suite of image enhancement techniques as a preprocessing step before applying Mask R-CNN. Enhancement methods such as Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), Contrast Limited Adaptive Histogram Equalization (CLAHE), and Blind Deblurring (BD) are employed to improve contrast and sharpness [16-22]. By integrating various preprocessing strategies with the instance segmentation capabilities of Mask R-CNN, this study aims to address the limitations identified in previous research. The contributions of this work include a comprehensive comparative evaluation of multiple image enhancement techniques on segmentation performance, the development of a customized preprocessing pipeline specifically designed for low-quality medical images, and the application of the proposed approach to fetal heart anatomical segmentation a domain that has received limited attention in the context of MIIS using enhanced Mask R-CNN frameworks. This integrated methodology is anticipated to improve the reliability and accuracy of instance segmentation in suboptimal imaging conditions, thereby supporting the advancement of MIIS applications in clinical settings.

2. RELATED WORKS

Research on Medical Image Instance Segmentation (MIIS) has advanced rapidly alongside developments in deep learning (DL) technologies, particularly Convolutional Neural Networks (CNNs). One of the most widely adopted DL-based approaches for medical image segmentation is Mask R-CNN, known for its ability to detect and delineate individual objects at the instance level with high precision. He et al. [23] introduced Mask R-CNN as an extension of Faster R-CNN, incorporating a parallel mask prediction branch that enables pixel-level object segmentation and classification. Wang et al. [23] applied Mask R-CNN to the segmentation of skin lesions in dermoscopic images, demonstrating significantly improved accuracy compared to conventional methods.

Various image enhancement techniques have been explored to enhance segmentation accuracy in medical imaging further. Saifullah and Dreżewski [24] proposed a CNN-based segmentation approach integrated with several preprocessing scenarios, including HE, CLAHE, and two hybrid methods: HE-CLAHE and CLAHE-HE. The evaluation was conducted on two publicly available radiological datasets: Lung CT-Scan and Chest X-ray. Experimental results indicated that the CLAHE-HE preprocessing scenario consistently improved segmentation performance, achieving a Dice Similarity Coefficient (DSC) of up to 0.92 and a Structural Similarity Index Measure (SSIM) of up to 0.97.

Several other studies have also combined Mask R-CNN with image enhancement techniques. Balasubramanian et al. [25] proposed a DL-based approach for liver tumor segmentation and classification from CT images. Their model consists of three stages: preprocessing using HE and a median filter, liver segmentation using an enhanced Mask R-CNN, and classification using APESTNet—an Enhanced Swin Transformer Network with Adversarial Propagation. The results demonstrated superior performance across various CT image types, with high efficiency and robustness to noise.

Khan et al. [26] developed a deep learning-based liver disease segmentation and classification system using a Customized Mask R-CNN (cm-RCNN). Preprocessing was performed using AHE to enhance image quality. Segmentation was conducted with cm-RCNN employing modified ReLU and sigmoid activation functions. Feature extraction involved texture, morphological, and deep features derived from ResNet and median binary pattern. The final classification stage utilized an ensemble approach combining SqueezeNet and DeepMaxout.

Table 1. Review of studies on the use of Mask R-CNN in medical imaging

Researcher (Year)	Research Object	Methodology	Preprocessing Techniques	Key Findings
He et al. (2018) [10]	COCO & Pascal VOC	Mask R-CNN	-	Introducing Mask R-CNN, accurate segmentation.
Wang et al. (2024) [23]	Dermatoscopic images of skin lesions	adjustments to Mask R-CNN parameters	-	Significant improvement in segmentation accuracy over conventional methods.
Saifullah and Dreżewski (2023) [24]	Lung CT-Scan and Chest X-ray images	CNN with preprocessing scenarios	HE, CLAHE, HE-CLAHE, CLAHE-HE	CLAHE-HE consistently improved performance with DSC up to 0.92 and SSIM up to 0.97.
Balasubramanian et al. (2023) [25]	CT images of liver tumors	Mask R-CNN + APESTNet (Enhanced Swin Transformer)	HE, Median filter	High accuracy and robustness to noise in liver tumor segmentation and classification.
Khan et al. (2024) [26]	Liver disease CT images	Customized Mask R-CNN + SqueezeNet + DeepMaxout	AHE	Accurate segmentation and classification using ensemble with handcrafted and deep features.
Han et al. (2023) [27]	SEM images of blurry nanoparticles	BL-Mask R-CNN + DeblurGAN-v2	Blind Deblurring	Accuracy improved from 0.8339 to 0.9613, demonstrating the effectiveness of deblurring preprocessing.
Nurmaini et al. (2021) [28]	Fetal ultrasound heart images	Mask R-CNN	-	Improved detection of septal defects such as ASD and VSD.
Sapitri et al. (2023) [29]	Real-time fetal sub-heart structure in ultrasound video	YOLO framework	-	Achieved 82.10% average precision and 17 FPS for real-time detection.

Han et al. [27] proposed BL-Mask R-CNN, an instance segmentation approach for blurred scanning electron microscope (SEM) images of nanoparticles. This method integrates DeblurGAN-v2 as a preprocessing stage to restore texture details and enhance image clarity prior to segmentation by Mask R-CNN. Evaluation on the NFFAEUROPE dataset revealed an increase in detection accuracy from 0.8339 to 0.9613, highlighting the effectiveness of deblurring preprocessing in improving segmentation performance on low-quality images.

In the context of fetal heart segmentation, Nurmaini et al. [28] developed a Mask R-CNN model to segment fetal heart structures in ultrasound images. The model showed improved accuracy in detecting septal anomalies such as atrial septal defect (ASD) and ventricular septal defect (VSD). Meanwhile, Iriani Sapitri et al. [29] proposed a real-time detection framework for substructures of the fetal heart in ultrasound video using the YOLO architecture, achieving an average precision of 82.10% and a processing speed of 17 frames per second. These studies collectively demonstrate that integrating instance segmentation techniques based on Mask R-CNN with appropriate preprocessing strategies is crucial for improving accuracy on challenging medical image datasets, including fetal heart anatomy. A comparative summary of these approaches is presented in Table 1.

3. PROPOSED METHOD

The methodology in this study comprises several interrelated stages, beginning with data acquisition and culminating in the evaluation of segmentation results. The overall workflow is illustrated in Figure 1.

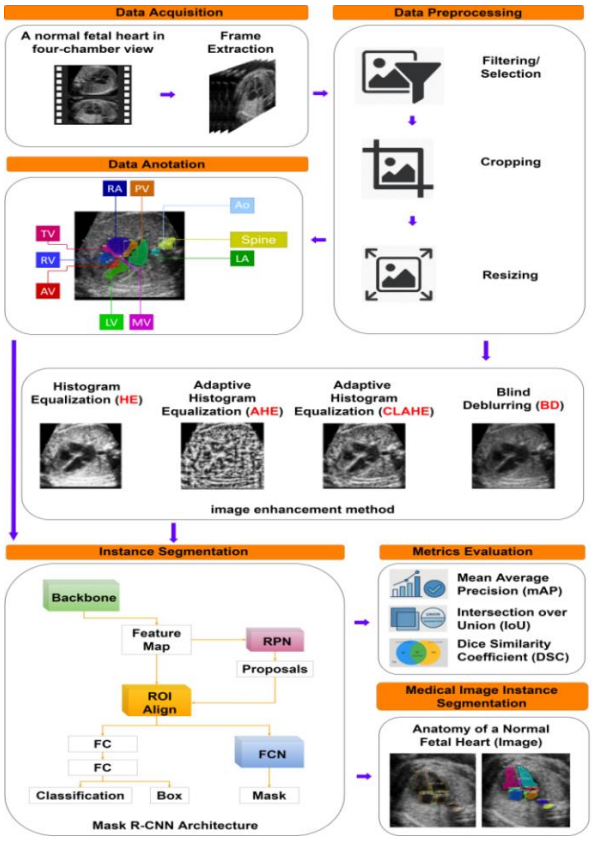


Figure 1. Workflow of fetal heart image segmentation using Mask R-CNN

3.1 Data acquisition

The initial stage in the SIIM process is data acquisition, as illustrated in Figure 1. The data used in this study were obtained from fetal echocardiography videos capturing the heart in the four-chamber view. These videos were sourced from an online platform with proper usage permission [30]. The video file was in .mp4 format, with a size of 13.7 MB, a duration of 178 seconds, and a frame rate of 30.00 fps. The entire video was then framed into two-dimensional ultrasound images, resulting in 357 images. These images include frames containing fetal heart objects some with one, two, or three fetal heart instances.

Additionally, some images did not contain fetal heart objects, or the heart objects were out of focus or blurred. Cropping was performed to ensure that the data used for the MIIS model met specific requirements for images containing multiple fetal heart objects or extraneous elements such as text. The output of the video extraction process and the resulting images are summarised in Table 2.

Table 2. Video extraction

Image Type	Dimensions	Number of Extracted Images
Images showing fetal heart objects	1280 × 720	114
Image showing multiple fetal heart objects	1280 × 720	50
Images showing fetal heart objects but out of focus	1280 × 720	105
Images not showing any fetal heart objects	1280 × 720	88
Total		357

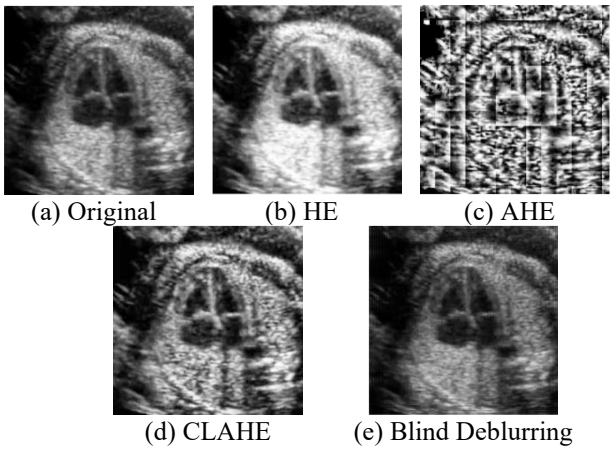
3.2 Data preprocessing

The data preprocessing stage includes selection, cropping, and resizing of images, aimed at producing a dataset that represents explicitly the fetal heart structure in normal conditions. As a result of this process, a dataset consisting of 176 ultrasound images of normal fetal hearts was compiled. This dataset was divided into two main subsets: 140 images were allocated for training, while 36 images were designated for validation. Each image in this dataset represents ten primary anatomical classes of the fetal heart: Left Atrium (LA), Right Atrium (RA), Left Ventricle (LV), Right Ventricle (RV), Tricuspid Valve (TV), Pulmonary Valve (PV), Mitral Valve (MV), Aortic Valve (AV), Aorta (Ao), and Spine. Consequently, the training data contains 1,400 anatomical class labels, while the validation data includes 360 class labels. Detailed information regarding the data distribution is presented in Table 3.

The subsequent stage involves applying image enhancement techniques to improve the visibility of cardiac structures and minimise the effects of blur or noise. Several methods are utilised in this process, including HE, which enhances the global contrast of the image; AHE, which adjusts the local contrast in different regions; and CLAHE, which limits excessive contrast enhancement in overly bright or dark areas. Additionally, BD reduces blur without requiring explicit knowledge of the blur kernel. The results of these enhancement techniques are presented in Figure 2.

Table 3. Dataset MIIS

Data	Number of Original Images	Number of HE Image Enhancements	Number of AHE Image Enhancements	Number of CLAHE Image Enhancements	Number of BD Image Enhancements
Training Data	140	140	140	140	140
Validation Data	36	36	36	36	36



(a) Original (b) HE (c) AHE (d) CLAHE (e) Blind Deblurring

Figure 2. Fetal heart image

Image quality enhancement aims to improve visual appearance to support more accurate analysis and feature extraction. In this study, several image enhancement methods are employed, including HE, as defined by Eq. (1) [19].

$$H(i) = \sum_{x=1}^{M} \sum_{y=1}^{N} \delta(f(x,y) = i) \tag{1}$$

AHE, as defined by Eq. (2) [20],

$$f'(x,y) = \frac{(L-1)}{|W|} \sum_{k=0}^{f(x,y)} H_{W_{x,y}}(k) \tag{2}$$

CLAHE, as defined by Eq. (3) [31],

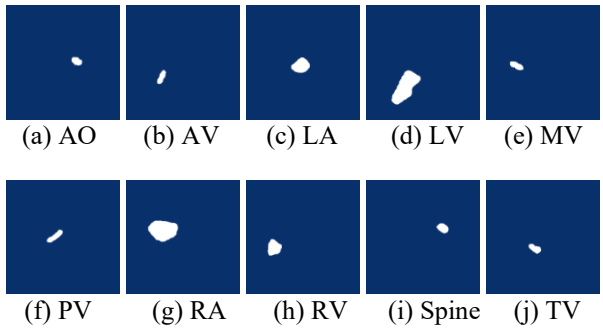
$$f'(x,y) = \frac{(L-1)}{|W|} \sum_{k=0}^{f(x,y)} \min\left(H_{W_{x,y}}(k), T\right) + \frac{E}{L} \tag{3}$$

and BD, as defined by Eq. (4) [22].

$$\hat{S} = \arg min_s \parallel I - \hat{K} * S \parallel^2 + \lambda R(S) \tag{4}$$

3.3 Data annotation

In parallel with the image enhancement process, all normal fetal heart images undergo annotation using ten labels that represent the anatomical features of the fetal heart. The annotation is carried out by precisely placing polygon points on the heart regions within each image. Once completed, the annotated images are exported in JavaScript Object Notation (JSON) format for further processing. In addition to enhancing image quality, this annotation process aims to generate ground truth images manually labelled data created by experts with the competence to identify anatomical features and structures in medical images. These ground truth images are the primary reference for training and evaluating image processing models. An example of the annotation result is shown in Figure 3.



(a) AO (b) AV (c) LA (d) LV (e) MV

(f) PV (g) RA (h) RV (i) Spine (j) TV

Figure 3. Ground truth of annotation results

3.4 Instance segmentation with Mask R-CNN

We employed the Mask R-CNN technique for MIIS [23]. Mask R-CNN features a core architecture that utilises a convolutional neural network (CNN) backbone, such as ResNet, to extract feature maps from the input images. These feature maps are then processed by a Region Proposal Network (RPN) to generate candidate bounding boxes that indicate potential object locations, enabling the analysis to focus on specific regions of interest.

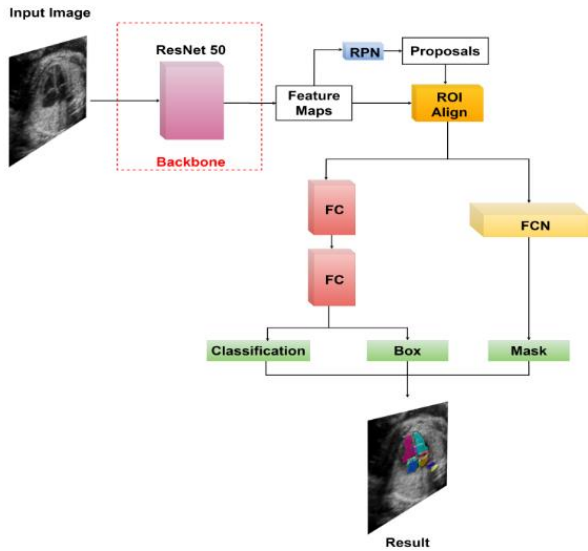


Figure 4. Mask R-CNN architecture for deep learning-based MIIS

The primary distinction between Mask R-CNN and its predecessor, Faster R-CNN, lies in using ROI Align instead of ROI Pooling. ROI Align is designed to preserve the spatial alignment of feature maps, which is crucial for improving the accuracy of detection and segmentation. This is especially important in medical imaging, where spatial distortion caused by ROI Pooling can significantly degrade the quality of the resulting segmentation [10]. The architecture and workflow of

Mask R-CNN as implemented in this study are illustrated in Figure 4.

3.5 Model evaluation

To evaluate the performance of image segmentation models, three primary metrics are commonly used: Mean Average Precision (mAP), Intersection over Union (IoU), and Dice Similarity Coefficient (DSC). Among these, mAP serves as the main evaluation metric for assessing segmentation accuracy, as it accounts for the average precision across various IoU thresholds and reflects the model's capability to detect and classify objects accurately. The mAP value is computed as the mean of the Average Precision (AP) scores across all segmented object classes, as defined in Eq. (5). [32]:

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

where, N is the number of classes and denotes the Average Precision for the i-th class.

Furthermore, IoU measures the extent of overlap between the predicted segmentation and the ground truth. The IoU value is calculated by comparing the area of intersection between the prediction and the ground truth with the area of their union, as defined in Eq. (6) [33]:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{6}$$

where, A represents the pixels in the predicted segment and B represents the pixels in the ground truth segment. A higher IoU value indicates a more accurate segmentation performed by the model.

Next, the Dice Similarity Coefficient (DSC), or Dice Score, is used to measure the similarity between two segments, specifically between the predicted segmentation and the ground truth. This metric is particularly useful in segmentation evaluation due to its higher sensitivity to small object sizes. The DSC is defined by Eq. (7) [34]:

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \tag{7}$$

where, A and B represent the pixels of the predicted and ground truth segments, respectively. The DSC value ranges from 0 to 1, with a value of 1 indicating a perfect segmentation.

3.6 Training setup

At this stage, the environment and training parameters for the Mask R-CNN model were configured to ensure an optimal and consistent training process. This phase aimed to develop an instance segmentation model capable of accurately identifying and distinguishing anatomical structures in fetal heart images, through system configuration and training parameter adjustments tailored to the characteristics of the medical imaging data. An ablation study was conducted to determine the optimal hyperparameter settings. Key parameters such as learning rate, batch size, and number of epochs were varied to assess their impact on model accuracy, convergence, and overall performance. The results indicated that the best balance between stability and efficiency was achieved with a learning rate of 0.01, a batch size of 1 due to GPU capacity limitations, and 20 epochs with 500 steps per epoch. The detailed training parameters are presented in Table 4.

Table 4. Model training setting and environment

Parameter	Details
Hardware	Nvidia GeForce GTX 1050 Ti GPU with 768 CUDA cores, a GPU clock of 1392/1506 MHz, 4GB of GDDR5 GPU memory, and a memory bandwidth of 112.1 GB/s
Train Environment	Python 3.6.13, with the TensorFlow 1.14.0 and Keras 2.3.1 libraries, and Protobuf 3.19.6
Batch Size	1
Epochs	20
Step Per Epoch	500
Learning Rate	0.01
Learning Momentum	[0.7,0.9]
Optimization Algorithm	Adam optimizer
Image Size	512×512
Image Enhancement	[Original, HE, AHE, CLAHE, Blind Deblurring]

3.7 Model segmentation design

In this study, image segmentation models were developed using uniformly controlled experimental parameters to ensure consistency and fairness in performance evaluation. The fixed parameters applied across all experiments included 500 steps per epoch, 20 training epochs, a ResNet-50 backbone, Stochastic Gradient Descent (SGD) as the optimizer, a batch size of 1, and a learning rate of 0.01. These parameter values were selected based on the default configuration of Mask R-CNN, while the batch size was adjusted to accommodate the limitations of the available GPU resources.

The independent variables in this experiment were the learning momentum (LM) values and the image enhancement techniques applied to the training data. Two momentum values, 0.7 and 0.9, were investigated. These values were chosen based on prior studies demonstrating their effectiveness in accelerating convergence and stabilizing SGD training in medical imaging contexts. A lower momentum (e.g., 0.7) allows greater sensitivity to gradient updates, while a higher momentum (e.g., 0.9) offers more stability and smoother optimization, making this range suitable for comparative performance analysis.

The main image enhancement techniques explored in this study include HE, AHE, CLAHE, and BD. In addition to evaluating each enhancement method individually, this study also examined potential synergies by applying various combinations, including HE + CLAHE, AHE + CLAHE, HE + BD, AHE + BD, HE + CLAHE + BD, and AHE + CLAHE + BD. These were also compared with no enhancement (i.e., the original images).

In total, 22 different models were trained and evaluated, representing all possible combinations of the eleven enhancement configurations with two different momentum values applied during the optimization process. This comprehensive experimental design was conceived to systematically assess the effects of both preprocessing strategies and momentum variations on instance segmentation accuracy, all under consistent training conditions. The complete list of model configurations is presented in Table 5.

Table 5. Model experimentation

Model	LM	Image Enhancement
model 1	0.7	Original
model 2	0.9	Original
model 3	0.7	HE
model 4	0.9	HE
model 5	0.7	AHE
model 6	0.9	AHE
model 7	0.7	CLAHE
model 8	0.9	CLAHE
model 9	0.7	Blind Deblurring
model 10	0.9	Blind Deblurring
model 11	0.7	HE + CLAHE
model 12	0.9	HE + CLAHE
model 13	0.7	AHE + CLAHE
model 14	0.9	AHE + CLAHE
model 15	0.7	HE + Blind Deblurring
model 16	0.9	HE + Blind Deblurring
model 17	0.7	AHE + Blind Deblurring
model 18	0.9	AHE + Blind Deblurring
model 19	0.7	HE + CLAHE + Blind Deblurring
model 20	0.9	HE + CLAHE + Blind Deblurring
model 21	0.7	AHE + CLAHE + Blind Deblurring
model 22	0.9	AHE + CLAHE + Blind Deblurring

4. RESULTS

This study aims to evaluate the performance of several Mask R-CNN models in object detection and instance segmentation tasks using fetal heart anatomical images.

4.1 Experimental results and loss analysis

A total of 22 Mask R-CNN models were evaluated based on twelve key loss components, encompassing all aspects of training and validation: mean loss, mean validation loss, mean RPN classification loss, mean validation RPN classification loss, mean RPN bounding box loss, mean validation RPN bounding box loss, mean Mask R-CNN classification loss, mean validation Mask R-CNN classification loss, mean Mask R-CNN bounding box loss, mean validation Mask R-CNN bounding box loss, mean Mask R-CNN mask loss, and mean validation Mask R-CNN mask loss. This evaluation aimed to assess the stability of the learning process and the generalization capability of the models to unseen data.

In general, the mean training loss ranged from 0.5509 to 0.6441, with the lowest value achieved by Model 9, indicating efficient learning during training. Meanwhile, the mean validation loss exhibited a wider range, from 3.6182 to 4.3479. Model 17 recorded the lowest validation loss, reflecting relatively better generalization ability, whereas Model 2 recorded the highest mean validation loss, indicating a strong tendency toward overfitting.

Regarding the RPN classification loss, the average training values were very small, ranging from 0.0074 (Model 15) to 0.0096 (Model 12), suggesting that the models generally succeeded in distinguishing between object and non-object regions during training. However, the validation values varied significantly, with Model 18 recording the highest value of 0.7328. This suggests that the model struggled to detect objects on unseen data. The RPN bounding box loss, which measures the accuracy of the RPN in predicting object proposal boxes, had its lowest training value in Model 9 at 0.0894, while the best validation result was achieved by Model 17 with 0.5794. Conversely, Model 18 again showed the highest value of 0.9604, indicating instability in localizing bounding boxes during validation.

The Mask R-CNN classification loss showed much higher values compared to other components. During training, the loss ranged from 0.1216 to 0.1338, but rose substantially during validation, from 1.4452 (Model 19) to 1.8048 (Model 14). This underscores that object classification remains a major challenge, particularly when models are tested on new data. Nevertheless, despite having the highest classification loss, Model 14 managed to achieve outstanding segmentation results. For the Mask R-CNN bounding box loss, the models remained relatively stable in both training and validation phases, with training values ranging from 0.1086 to 0.1409 and validation values from 0.3942 to 0.4625. Model 11 and Model 17 demonstrated the most optimal validation bounding box loss, indicating consistent spatial accuracy on unseen data.

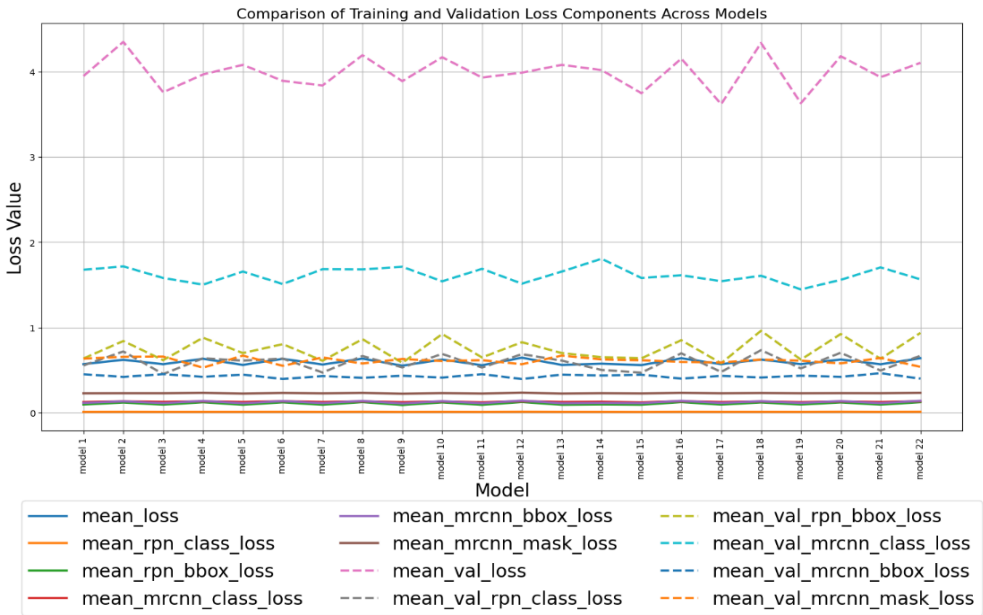


Figure 5. Comparison of training and validation loss components across models

Finally, the Mask R-CNN mask loss, which is closely related to pixel-level segmentation quality, ranged from 0.2217 to 0.2350 during training and from 0.5277 to 0.6688 during validation. Model 4 achieved the lowest validation mask loss (0.5277), while Models 5 and 13 recorded the highest values (0.6688). This suggests that using AHE as the sole image enhancement technique may degrade segmentation performance during validation, possibly due to excessive contrast enhancement introducing noise to pixel features.

Overall, this analysis indicates that the combination of image enhancement techniques and training parameters significantly influences each loss component. Models 9 and 17 consistently demonstrated a balanced performance between training and validation, particularly in the RPN and bounding box components. Meanwhile, despite having the highest classification loss on validation, Model 14 produced superior segmentation performance, suggesting that high loss in a single component does not necessarily correlate negatively with final segmentation quality. Figure 5 presents a comprehensive comparison of all training and validation loss components across the 22 evaluated models. This visualization reveals consistent trends, where most models exhibit higher validation losses than training losses a common indication of overfitting. The graph also highlights stark contrasts between components such as mrcnn_class_loss and mrcnn_mask_loss, which can help identify specific weaknesses in each model.

4.2 Comparison of evaluation metrics for segmentation models

The performance evaluation of medical image segmentation models was carried out using three key metrics: Intersection over Union (IoU), Dice Similarity Coefficient (DCS), and mean Average Precision (mAP). These metrics provide a comprehensive assessment of both the spatial accuracy and detection precision of anatomical structure segmentation in ultrasound images (Figure 6).

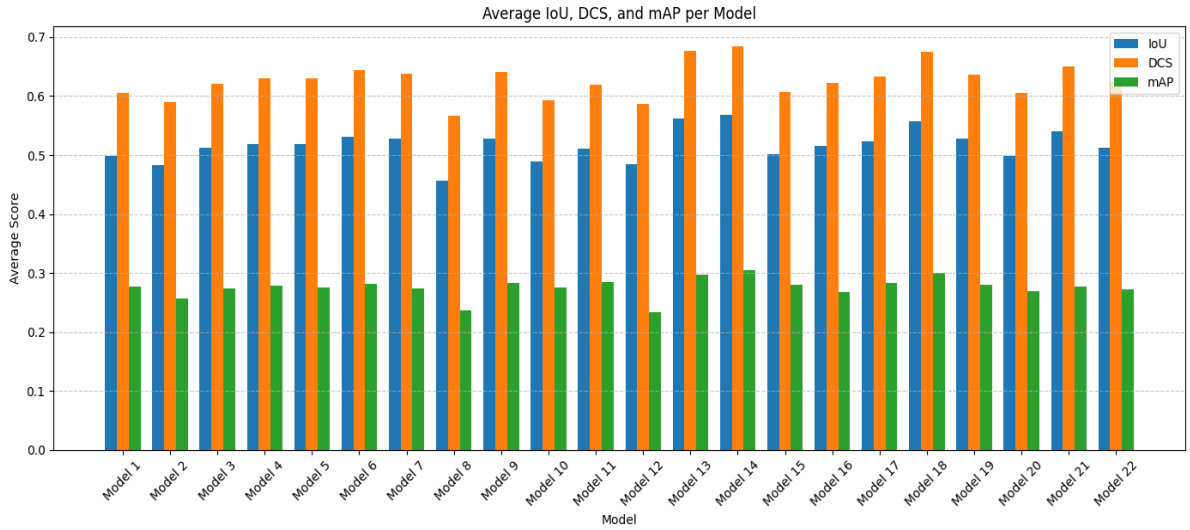
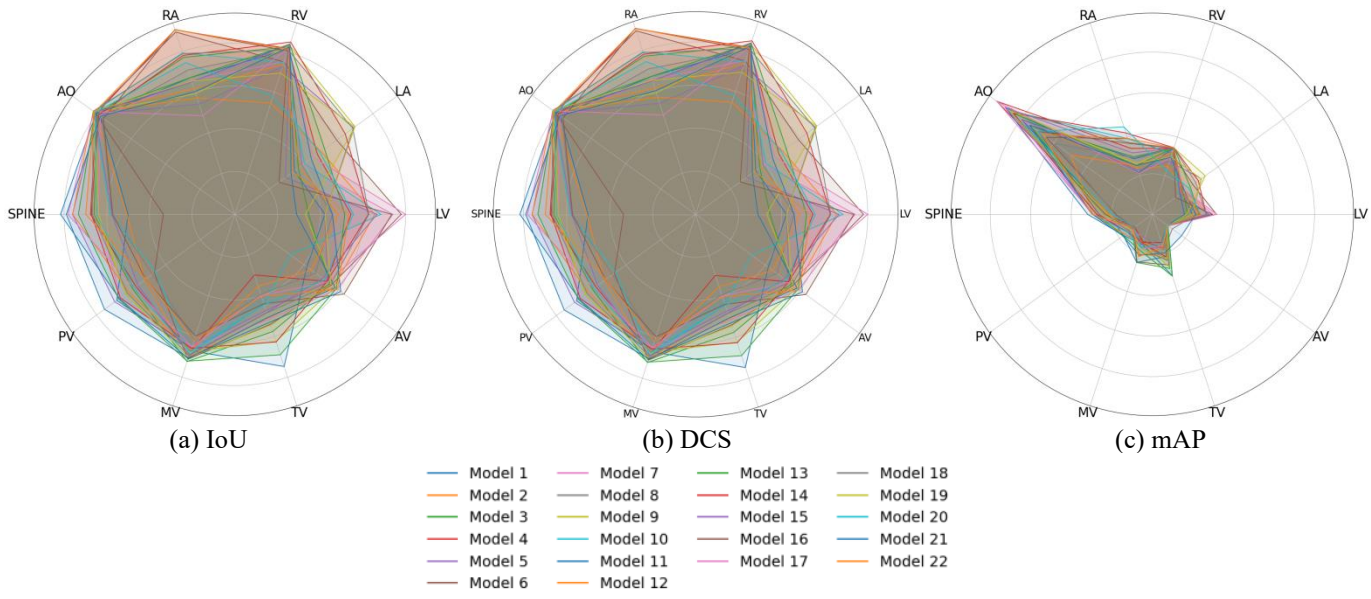


Figure 6. Average IoU, DCS, and mAP per model



(a) IoU (b) DCS (c) mAP

(d) legend

Figure 7. Comparison of IoU, DCS, and mAP per class across 22 Mask R-CNN models

An analysis of the 22 evaluated models revealed notable variation in performance metrics, which depended on the applied image enhancement technique and momentum parameter. The average IoU ranged from 0.4473 (Model 8) to 0.5887 (Model 14), with other strong performers including Model 13 (0.5744), Model 18 (0.5800), and Model 5 (0.5233). A similar trend was observed for DCS, where scores varied from 0.5765 (Model 11) to 0.7032 (Model 14), followed by Model 13 (0.6819) and Model 18 (0.6933). For mAP, Model 14 again achieved the highest score (0.3049), while Model 13 (0.2965) and Model 18 (0.3007) also demonstrated superior performance.

To further assess the consistency and reliability of the models, the standard deviation for each evaluation metric was computed across all models. The resulting values were 0.5120 ± 0.0366 for IoU, 0.6145 ± 0.0382 for DCS, and 0.2753 ± 0.0184 for mAP. These relatively low deviations suggest a moderate level of variability, indicating that while preprocessing strategies significantly affect segmentation quality, most configurations produced stable outcomes. Notably, Model 14 not only achieved the highest performance across all metrics but also demonstrated statistical robustness, reinforcing its effectiveness for segmenting small and low-contrast anatomical structures in fetal heart ultrasound images. This superior performance can be attributed to the synergistic use of AHE and CLAHE. AHE enhances local contrast in small regions, helping to delineate fine anatomical boundaries, while CLAHE limits noise amplification by clipping histogram peaks, making it suitable for noisy medical images. Their combined application improves feature visibility without over-enhancing artifacts. Additionally, the use of a higher momentum value (0.9) facilitates more stable and consistent learning by reducing the variance in gradient updates, which supports better generalization to unseen data. Together, these factors contribute to the strong segmentation performance observed in Model 14.

The evaluation was conducted based on ten fetal heart anatomical classes: LV, LA, RV, RA, AO, Spine, PV, MV, TV, and AV. Figure 7 illustrates performance comparisons across classes. For the LV class, segmentation performance varied across models, with Model 17 achieving the best results (IoU: 0.7169, DCS: 0.8005, and mAP: 0.3182). In the LA class, peak performance was observed in Models 5 and 9, with Model 5 recording the highest DCS (0.6878) and Model 9 the highest IoU (0.5970). Model 14, however, achieved the top mAP (0.2810). For the RV class, Model 14 again excelled with an IoU of 0.7365, DCS of 0.8468, and the highest mAP of 0.3451. In the RA class, Models 16 and 22 stood out for IoU and DCS, respectively, while Model 20 achieved the highest mAP (0.4534). The AO class showed consistent performance, with Model 4 reaching the highest mAP (0.9494), while Models 18 and 12 excelled in IoU and DCS. For the Spine class, Models 11 and 5 recorded the best DCS (0.8168) and IoU (0.6176), respectively, with Model 14 maintaining stable performance (mAP: 0.2745). In the PV class, Model 21 achieved the highest DCS (0.7550), and Model 13 had the top mAP (0.1781). For MV, Model 13 again led in mAP (0.1892), while Models 3 and 6 had the highest DCS and IoU. In the TV class, Model 13 showed the best performance across all metrics. Finally, in the AV class, Models 21 and 18 led in DCS, while Model 13 achieved the best mAP (0.0942).

5. DISCUSSIONS AND FINDINGS

Evaluation based on average IoU reveals considerable performance variation across models. Model 14 recorded the highest average IoU of 0.5887, indicating superior segmentation capability across all classes. Specifically, this model also demonstrated strong performance in segmenting the RV, RA, and AO structures, critical components in fetal heart imaging. Models 6 and 18 also showed competitive performance, with average IoU scores above 0.50, suggesting stable segmentation across various anatomical structures. In contrast, models such as Model 1 and Model 11 exhibited relatively low performance, particularly in classes like LA and AV, which may be attributed to their small size or low contrast in the original images.

The measurement results with DCS support the findings from the IoU metrics, Model 14 again ranked the highest (average DCS = 0.7032), reinforcing the conclusion that this model can produce consistent segmentation with high overlap with the ground truth. Models 18 and 13 also demonstrated strong DCS performance, with average scores above 0.68. Classes such as RV and AO generally exhibited high DCS scores across most models, likely due to their clearer morphological contours and boundaries. In contrast, classes like TV and AV tended to have lower DCS scores, indicating that these structures are more challenging to segment accurately.

In the mAP evaluation, Model 14 again demonstrated the best performance with an average score of 0.3049, reinforcing its dominance across various evaluation metrics. This indicates that the model is not only capable of producing accurate segmentation predictions but also maintains consistency across different instances. Interestingly, although some models achieved relatively high IoU or DCS values, they did not necessarily maintain similarly high performance in mAP. For example, Model 6, which recorded strong IoU and DCS scores, showed a slightly lower mAP compared to models 14 and 13. This suggests that strong spatial segmentation does not always guarantee precise instance-level detection.

Among all evaluated models, Model 14 consistently demonstrated the best performance across the three main evaluation metrics: average IoU (0.5887), DCS (0.7032), and mAP (0.3049). The success of this model can be attributed to two key factors: the use of a LM value of 0.9 and the application of a combined preprocessing technique using AHE and CLAHE. The combination of AHE and CLAHE proved effective in enhancing local contrast in fetal cardiac ultrasound images, which is crucial for highlighting boundaries of small and low-contrast structures such as AV, TV, and LA. Meanwhile, the relatively high LM value helped the model achieve stability during training while maintaining good generalisation on test data. These findings indicate that selecting appropriate image preprocessing strategies and tuning hyperparameters optimally significantly impact segmentation performance.

6. CONCLUSION

This study highlights the critical role of image enhancement and training parameter optimization in improving deep learning-based segmentation of fetal heart structures in ultrasound images. Specifically, the combination of AHE and

CLAHE was found to be particularly effective for enhancing local contrast, which in turn improved the delineation of small and low-contrast anatomical regions. Models trained with higher momentum values exhibited better generalization and stability, suggesting the importance of careful hyperparameter tuning. Overall, these findings underscore that successful segmentation of complex anatomical structures requires not only robust model architectures but also thoughtfully designed preprocessing strategies.

Future work will focus on extending this model to handle 3D ultrasound volumes and incorporating post-processing techniques to refine segmentation boundaries. Additionally, ensemble-based architectures may be explored to dynamically select or combine models based on anatomical class characteristics, with the goal of enhancing robustness and enabling practical clinical deployment in fetal cardiac assessment.

ACKNOWLEDGMENT

We would like to thank the University of South Sumatra and the Institute for Research and Community Service (LPPM) for their invaluable support during the preparation of this study. We also express our appreciation to Dr. Anthony L. Filly for granting permission to use his video data.

REFERENCES

[1] Arabahmadi, M., Farahbakhsh, R., Rezazadeh, J. (2022). Deep learning for smart healthcare—A survey on brain tumor detection from medical imaging. Sensors, 22(5): 1960. https://doi.org/10.3390/s22051960

[2] Shurrab, S., Duwairi, R. (2022). Self-supervised learning methods and applications in medical imaging analysis: A survey. PeerJ Computer Science, 8: e1045. https://doi.org/10.7717/peerj-cs.1045

[3] Gichoya, J.W., Banerjee, I., Bhimireddy, A.R., Burns, J.L., et al. (2022). AI recognition of patient race in medical imaging: A modelling study. The Lancet Digital Health, 4(6): e406-e414. https://doi.org/10.1016/S2589-7500(22)00063-2

[4] Oulefki, A., Agaian, S., Trongtirakul, T., Laouar, A.K. (2021). Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images. Pattern Recognition, 114: 107747. https://doi.org/10.1016/j.patcog.2020.107747

[5] Niyas, S., Pawan, S.J., Kumar, M.A., Rajan, J. (2022). Medical image segmentation with 3D convolutional neural networks: A survey. Neurocomputing, 493: 397-413. https://doi.org/10.1016/j.neucom.2022.04.065

[6] Wang, R.S., Lei, T., Cui, R.X., Zhang, B.T., Meng, H.Y., Nandi, A.K. (2022). Medical image segmentation using deep learning: A survey. IET Image Processing, 16(5): 1243-1267. https://doi.org/10.1049/ipr2.12419

[7] Hermessi, H., Mourali, O., Zagrouba, E. (2021). Multimodal medical image fusion review: Theoretical background and recent advances. Signal Processing, 183: 108036. https://doi.org/10.1016/j.sigpro.2021.108036

[8] Hu, M.D., Zhong, Y., Xie, S.X., Lv, H.B., Lv, Z.H. (2021). Fuzzy system based medical image processing for brain disease prediction. Frontiers in Neuroscience, 15: 714318. https://doi.org/10.3389/fnins.2021.714318

[9] Ramesh, K.K.D., Kumar, G.K., Swapna, K., Datta, D., Rajest, S.S. (2021). A review of medical image segmentation algorithms. EAI Endorsed Transactions on Pervasive Health & Technology, 7(27): e6. http://doi.org/10.4108/eai.12-4-2021.169184

[10] He, K., Gkioxari, G., Dollár, P., Girshick, R. (2017). Mask R-CNN. In 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. pp. 2980-2988. http://doi.org/10.1109/ICCV.2017.322

[11] Hesamian, M.H., Jia, W.J., He, X.J., Kennedy, P. (2019). Deep learning techniques for medical image segmentation: Achievements and challenges. Journal of digital imaging, 32: 582-596. http://doi.org/10.1007/s10278-019-00227-x

[12] Sirisha, A., Chaitanya, K., Krishna, K.V.S.S.R., Kanumalli, S.S. (2021). Intrusion detection models using supervised and unsupervised algorithms - a comparative estimation. International Journal of Safety and Security Engineering, 11(1): 51-58. https://doi.org/10.18280/ijsse.110106

[13] Ullah, I., Ali, F., Shah, B., El-Sappagh, S., Abuhmed, T., Park, S.H. (2023). A deep learning based dual encoder–decoder framework for anatomical structure segmentation in chest X-ray images. Scientific Reports, 13(1): 791. https://doi.org/10.1038/s41598-023-27815-w

[14] Gulakala, R., Markert, B., Stoffel, M. (2023). Rapid diagnosis of Covid-19 infections by a progressively growing GAN and CNN optimisation. Computer Methods and Programs in Biomedicine, 229: 107262. https://doi.org/10.1016/j.cmpb.2022.107262

[15] Elaraby, A., Elansary, I. (2021). A framework for multi-threshold image segmentation of low contrast medical images. Traitement du Signal, 3(2): 309-314. https://doi.org/10.18280/ts.380207

[16] Rao, G.S., Srikrishna, A. (2021). Image pixel contrast enhancement using enhanced multi histogram equalization method. Ingénierie des Systèmes d'Information, 26(1): 95-101. https://doi.org/10.18280/isi.260110

[17] Babu, V.S., Ram, N.V. (2020). Deep residual CNN with contrast limited adaptive histogram equalization for weed detection in soybean crops. Traitement du Signal, 39(2): 717-722. https://doi.org/10.18280/ts.390236

[18] Mohammed, M.H., Khalaf, N.A., Kareem, H.H., Daway, H.G. (2024). Enhancement of very low light images using the YIQ space based on the CLAHE and sigmoid mapping with high colour restoration. Revue d'Intelligence Artificielle, 38(2): 655-660. https://doi.org/10.18280/ria.380229

[19] AL-Azzawi, Z.H.N., Mahdi, W.H., Ahmed, S.K., Yasin, W.S. (2023). Medical image enhancement using histogram equalization techniques. AIP Conference Proceedings, 2591: 030036. https://doi.org/10.1063/5.0119326

[20] Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., Romeny, B.T.H., Zimmerman, J.B., Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. Computer vision, graphics, and image processing, 39(3): 355-368. https://doi.org/10.1016/S0734-189X(87)80186-X

[21] Balakrishnan, A.A., Dhanya, P.R., Anilkumar, S., Supriya, M.H. (2024). A novel L-CLAHE-based

intensification filter for enhancement of underwater images and pipeline tracking. IETE Journal of Research, 70(5): 4692-4701. https://doi.org/10.1080/03772063.2023.2225476

[22] Nourildean, S.W. (2024). Blind and non-blind deconvolution-based image deblurring techniques for blurred and noisy image. Tikrit Journal of Engineering Sciences, 31(1): 12-22. https://doi.org/10.25130/tjes.31.1.2

[23] Wang, Z., Wang, C., Peng, L., Lin, K.B., Xue, Y., Chen, X., Bao, L.L., Liu, C., Zhang, J.L., Xie, Y. (2024). Radiomic and deep learning analysis of dermoscopic images for skin lesion pattern decoding. Scientific Reports, 14: 19781. http://doi.org/10.1038/s41598-024-70231-x

[24] Saifullah, S., Drężewski, R. (2023). Modified histogram equalization for improved CNN medical image segmentation. Procedia Computer Science, 225: 3021-3030. http://doi.org/10.1016/j.procs.2023.10.295

[25] Balasubramanian, P.K., Lai, W.C., Seng, G.H., Selvaraj, J. (2023). APESTNet with Mask R-CNN for liver tumor segmentation and classification. Cancers, 15(2): 330. http://doi.org/10.3390/cancers15020330

[26] Khan, R., Su, L.L., Zaman, A., Hassan, H., Kang, Y., Huang, B.D. (2024). Customized m-RCNN and hybrid deep classifier for liver cancer segmentation and classification. Heliyon, 10(10): e30528. http://doi.org/10.1016/j.heliyon.2024.e30528

[27] Han, Y.H., Chen, S.J., Zhang, X.Y. (2023). Enhanced Mask R-CNN blur instance segmentation based on generated adversarial network. In Second International Conference on Electronic Information Technology (EIT 2023), Wuhan, China, Vol. 12719, pp. 1005-1012. http://doi.org/10.1117/12.2685510

[28] Nurmaini, S., Rachmatullah, M.N., Sapitri, A.I., Darmawahyuni, A., Tutuko, B., Firdaus, F., Partan, R.U., Bernolian, N. (2021). Deep learning-based computer-aided fetal echocardiography: Application to heart standard view segmentation for congenital heart defects detection. Sensors, 21(23): 8007. http://doi.org/10.3390/s21238007

[29] Sapitri, A.I., Nurmaini, S., Rachmatullah, M.N., Tutuko, B., Darmawahyuni, A., Firdaus, F., Rini, D.P., Islami, A. (2023). Deep learning-based real time detection for cardiac objects with fetal ultrasound video. Informatics in Medicine Unlocked, 36: 101150. http://doi.org/10.1016/j.imu.2022.101150

[30] My minifellowship. (2015). Mastering the fetal heart: Step 1. https://www.youtube.com/watch?v=xpxJFpORFmo.

[31] Hayati, M., Muchtar, K., Maulina, N., Syamsuddin, I., Elwirehardja, G.N., Pardamean, B. (2023). Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning. Procedia Computer Science, 216: 57-66. http://doi.org/10.1016/j.procs.2022.12.111

[32] Wang, B.N. (2022). A parallel implementation of computing mean average precision. arXiv preprint arXiv:2206.09504. https://doi.org/10.48550/arXiv.2206.09504

[33] Rezatofighi, H., Tsoi, T., Gwak, J., Sadeghian, A., Reid, I., Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. arXiv preprint arXiv:1902.09630. http://doi.org/10.48550/arXiv.1902.09630

[34] Wong, Y.M., Yeap, P.L., Ong, A.L.K., Tuan, J.K.L., Lew, W.S., Lee, J.C.L., Tan, H.Q. (2024). Machine learning prediction of Dice similarity coefficient for validation of deformable image registration. Intelligence-Based Medicine, 10: 100163. http://doi.org/10.1016/j.ibmed.2024.100163