Multi-scale input reconstruction network and one-stage instance segmentation for enhancing heart defect prediction rate

Sutarno[1], Siti Nurmaini[1], Ade Iriani Sapitri[1], Muhammad Naufal Rachmatullah[1], Bambang Tutuko[2], Annisa Darmawahyuni[1], Firdaus[1], Anggun Islami[1], Samsuryadi[3]
[1]Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, South Sumatera, Indonesia
[2]Faculty of Computer Science, Universitas Sriwijaya, South Sumatera, Indonesia
[3]Department of Informatic Engineering, Faculty of Computer Science, Universitas Sriwijaya, South Sumatera, Indonesia

Article Info

Article history:

Received Sep 15, 2023
Revised Oct 9, 2023
Accepted Nov 7, 2023

Keywords:

Congenital septal defect
Fetal heart
One stage instance
Segmentation
Real time
Ultrasound image enhancement

ABSTRACT

Artifacts and unpredictable fetal movements can hinder clear fetal heart imaging during ultrasound (US) scans, complicating anatomical identification. This study presents a new medical imaging approach that combines one-stage instance segmentation with US video enhancement for precise fetal heart defect detection. This innovation allows real-time identification and timely medical intervention. The study acquired 100 fetal heart US videos from an Indonesian Hospital featuring cardiac septal defects, generating 1,000 frames for training, validation, and testing. Utilizing a combination of the multi-scale input reconstruction network (MIRNet) for image enhancement and YOLOv8l-seg for real-time instance segmentation, the method achieved outstanding validation results, boasting a 99.50% mAP for bounding box prediction and 98.40% for mask prediction. It delivered a remarkable real-time processing speed of 68.4 frames per second. In application to new patients, the method yielded a 65.93% mAP for bounding box prediction and 57.66% for mask prediction. This proposed approach offers a promising solution to early fetal heart defect detection using US, holding substantial potential for enhancing healthcare outcomes.

Corresponding Author:

Siti Nurmaini
Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya
South Sumatera, Indonesia
Email: siti_nurmaini@unsri.ac.id

1. INTRODUCTION

Ultrasound (US) is a widely employed imaging technique in the medical field, serving as the primary diagnostic tool for a wide range of clinical scenarios. Despite advancements in US technologies and the well-established digital health systems supporting US obstetric imaging, the operator-dependent nature and manual operation of the procedure continue to present diagnostic challenges [1]. Furthermore, US obstetric scans may fail to provide sufficient information about the fetal heart due to the presence of various disturbances, such as signal dropout, attenuation, speckle, and acoustic shadows [1], [2]. The artifacts can significantly compromise the quality and reliability of fetal heart images obtained during US scans. These factors resulting in suboptimal anatomic fetal heart delineation [2], [3]. High-quality US imaging is critical for accurate diagnosis during obstetric scans, it is essential to prioritize achieving optimal image quality [3]. This is not only ensure the accuracy of the diagnosis but can also reduce the need for repeat scans and potential complications.

Obtaining high-quality obstetric scans in pregnancy women has long been a challenging issue [2]–[4]. With the increasing prevalence of obesity in this population, this challenge has become even more pressing [4]. The rate of obtaining inadequate images for assessing fetal heart defects in obese patients was significantly higher compared to a normal weight population [5]. The study found that while the rate was 6.4% for normal weight, it increased to 17.4% in obese patients where the fetus had a defect [5]. This underscores the importance of addressing the challenges of obtaining high-quality US images to ensure accurate diagnoses and appropriate medical management. Based on previous research, it has been established that improving the quality of images can greatly enhance the accuracy and efficiency of diagnostic procedures [6]. Consequently, the implementation of image enhancement algorithms in fetal US heart imaging can potentially improve the accuracy of diagnosing fetal heart defects.

With the rapidly growing image content, there is a pressing need to develop effective image enhancement algorithms [6]. Histogram equalization is the most commonly used approach. However, these networks are less effective in encoding contextual information and it frequently produces under- or over-enhanced images [7]. Prior to the deep learning (DL) era, numerous image enhancement algorithms have been proposed. Several enhancement algorithms mimicking human vision have been proposed in the literature [8]–[14]. Recently, convolutional neural networks (CNNs) have been successfully applied to general, as well as low-light, image enhancement problems [8], [9]. Notable works employ retinex-inspired networks [9], encoder-decoder networks [12]–[14], and generative adversarial networks (GANs) [10], [11]. As the fetal heart is small in size and US images can often contain noise and have low light quality, sometimes appearing dark [15], [16], it is important to choose an image enhancement algorithm that is suitable for these conditions. The multi-scale information representation network (MIRNet) algorithm, which proposes a combination of processes including image restoration, image denoising, and super-resolution, has the potential to greatly improve the visual quality and diagnostic accuracy of fetal heart US images [17]. With its ability to improve image resolution and reduce noise levels [17], we strongly believe such algorithm has the potential to significantly enhance the visual quality and diagnostic accuracy of fetal heart defects.

Heart defects are a group of congenital cardiac malformations that can be detected through US screening during pregnancy [18]. Due to their complexity, accurate diagnosis and appropriate treatment require the expertise of well-trained professionals [19], [20]. Routine clinical practice assessments involve manual segmentation of region of interests (RoIs), which is laborious and time consuming [19], [20]. The automatic image segmentation is tremendous research efforts were invested in developing novel methods by employing DL techniques [21]. Semantic segmentation is a first approach for visual scene understanding and focuses on classifying each pixel into a set of object classes [21], [22]. However, instance segmentation is a more challenging task because the goal of instance segmentation is to detect and segment each object instance found in the image [21]. The state-of-the-art approaches on instance segmentation are usually divided into two-stage detectors, e.g., mask region-based convolutional neural network (R-CNN) [21], and one-stage detectors, such as YOLO [23], [24], and YOLACT [25]. The two-stage instance segmentation approach consists of first generating a set of candidate RoIs and then segmenting and classifying the RoIs [21]. The two stages are generally applied sequentially, and therefore, these methods have difficulties in achieving real-time performance [20], [21]. Whereas, the one-stage instance segmentation approach focuses on directly generating an explicit localization [23]–[25]. To the best of our knowledge, we are the first to implement fetal heart US video enhancement, resulting in a robust and accurate real-time fetal heart defect prediction. The contributions of our study are as follows: i) introducing an innovative and unparalleled real-time one-stage instance segmentation approach for fetal heart defect detection; ii) developing an advanced US video enhancement technique utilizing a multi-scale input reconstruction network (MiRNet); and iii) conducting an exhaustive model evaluation solely based on previously unseen US images.

2. MATERIALS AND METHOD

2.1. Data preparation

Data preparation for one-stage instance segmentation models involves a series of steps to prepare the dataset for training and ensure that the model can learn to accurately predict object bounding boxes and segmentation masks from input images. Collecting a large dataset of US images with corresponding annotations of fetal heart object bounding boxes and segmentation masks is the first step in building a one-stage instance segmentation model. We collected 100 fetal heart US videos from Dr. Muhammad Hoesin General Hospital in Indonesia, which included cases of ASD, VSD, and AVSD. These videos were converted into a total of 1,000 images. The training process involved 650 images, the validation process used 200 images, and the remaining 150 images were used for testing as unseen data. To create the ground truth, two medical expert carefully outlines the object of interest in the US image using specialized software, named LabelMe. The resulting binary mask and bounding box is red and white image where the white pixels correspond to the object of interest as shown in Figure 1. The ground truth is an essential component of the segmentation model because it allows the model to learn which parts of the image are relevant to the object of interest and which are not. By comparing the model’s output with the ground truth, we can evaluate how well the model is performing and make improvements as necessary.

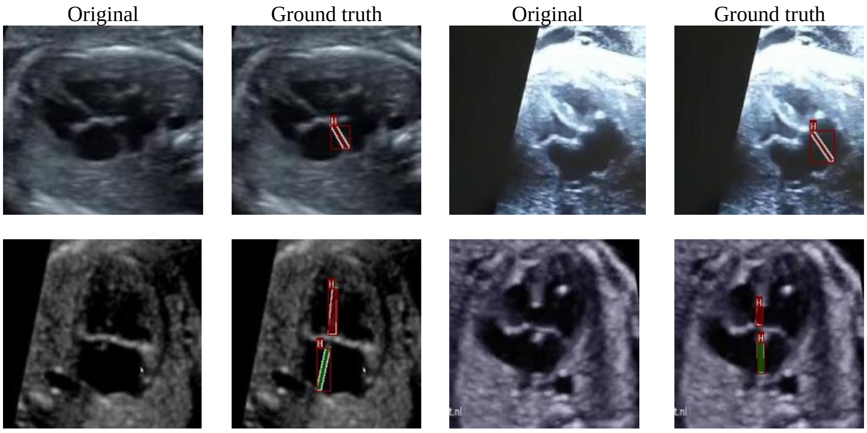


Figure 1. The sample of US image ground truth for segmentation model

2.2. The proposed model

In this section, we present an overview of the proposed one stage instance segmentation based on YOLOv8l, illustrated in Figure 2. Such algorithm employs the Darknet-53 architecture with 53 layers [23], which are organized into 5 blocks. The input size for the model is 416×416 pixels. Each convolutional layer uses a 3×3 kernel size and is followed by batch normalization and the LeakyReLU activation function. Additionally, max pooling with a 2×2 kernel and a stride of 2 is applied after each block. Down sampling is performed using convolutional layers with a stride of 2, while padding is used to maintain the same spatial dimensions in the output of each convolutional layer as the input. The final layer is a global average pooling layer that generates a feature vector used for object detection or classification. This is then followed by a fully connected layer with a SoftMax activation function for classification tasks. To accurately detect fetal heart defect condition, it is important to improve the quality of the US image.

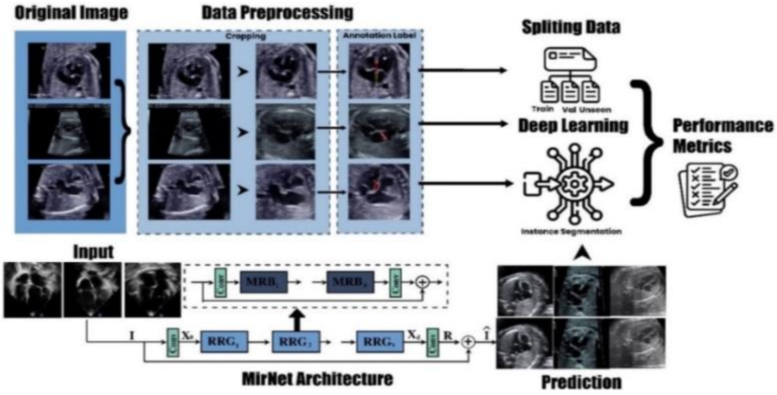


Figure 2. The research methodology of YOLOv8l-seg and MIRNet enhancement method

In this study, we implemented a simple image enhancement algorithm is MiRNet algorithm: one for feature extraction from the input image and the other for reconstructing the output image from the extracted features [17]. To generate the image enhancement model, we used 2263 high-resolution US images of infant hearts from our custom dataset. The MiRNet model was developed using this dataset, and we decreased the image quality by injecting four types of noise including Gaussian, poisson, salt and pepper, and speckle noises. The network consists of a series of mirrored residual blocks, which use residual connections to preserve low-level details while allowing for the network to learn high-level features as shown in Figure 2. The MiRNet architecture consists of 6 layers, comprising 5 residual network (ResNet) blocks and a final convolutional layer. Each layer in MIRNet comprises 64 filters, and each ResNet block consists of two convolutional layers, followed by batch normalization and rectified linear unit (ReLU) activation. The network is trained using the mean squared error (MSE) loss function with a learning rate of 0.0001. The Adam optimizer is used to optimize the weights of the MiRNet architecture.

2.3. Model evaluation

To assess the performance of the combined one-stage image segmentation and image enhancement, these essential metrics can be computed on either a validation set or a test set, which includes: i) intersection over union (IoU), to measure the overlap between the predicted segmentation mask and the ground truth mask [21], [23]. The IoU score ranges from 0 to 1, with a higher score indicating better performance; ii) mean average precision (mAP) to measure the accuracy of the model in localizing objects and predicting their class labels [21], [23]; iii) peak signal-to-noise ratio (PSNR) is used to measure the similarity between the original and restored images in terms of pixel-wise error. Higher PSNR values indicate better image quality for the image enhancement [17]; and iv) structural similarity index (SSIM) metric to measure the structural similarity between the original and restored images, taking into account factors such as luminance, contrast, and structure. Higher SSIM values indicate better image quality [17]. The proposed networks were implemented using Python and the PyTorch 1.7.1 library, and were trained on a computer with the following specifications: as a processor, an Intel® CoreTM i9-9920X CPU @ 3.50 GHz, 490,191 MB RAM, GeForce 2080 RTX Ti, made by the NVIDIA Corporation GV102 (rev a1); the operating system was Ubuntu 18.04.5 LTS.

3. RESULTS

YOLOv7-seg and YOLOv8l-seg architecture are both based on a modified version of the DarkNet architecture for real tine one stage segmentation, while YOLACT uses a ResNet-based backbone and a feature pyramid network for instance segmentation. However, YOLOv7-seg and YOLOv8l-seg use a panoptic feature fusion technique to perform instance segmentation [23]. They combines information from the object detection and semantic segmentation branches of the network to generate high-quality instance masks for each detected object. Our experiment compared the performance of three object real time instance segmentation models: YOLOv7-seg, YOLOv8l-seg, and YOLACT. We found that YOLOv8l-seg was the fastest and most accurate of the three models, followed by YOLOv7-seg and then YOLACT as shown in Table 1. The reason for this difference in performance is that YOLOv8l-seg uses a bounding box-based approach to detect and segment objects, which provides accurate results, while YOLACT uses a mask-based approach that can segment individual objects but may not be as precise as the bounding box approach. It mean YOLOv8l-seg is the best choice for real time segmentation like fetal heart US video tasks that require both speed and accuracy.

Table 1. Model comparison for robust segmentation: YOLACT, YOLACT ++, YOLOv7-seg, and YOLOv8l-seg

Backbone	Epoch	LR	Batch	mAP (%)	
			Size	B_box	Mask
YOLACT ResNet101	250	-	8	95.81	92.30
YOLACT ResNet50				97.38	91.46
YOLACT DarkNet53				92.21	91.83
YOLACT ResNet101	250	0.001	8	96.08	96.08
YOLACT ResNet50				92.91	91.04
YOLACT DarkNet53				96.46	94.84
YOLACT ResNet101	500	0.001	8	98.61	93.44
YOLACT ResNet50				94.19	87.77
YOLACT DarkNet53				96.05	91.24
YOLACT ResNet101	250	0.0001	8	96.09	91.46
YOLACT ResNet50				93.34	88.72
YOLACT DarkNet53				97.45	96.21
YOLACT ResNet101	500	0.0001	8	96.41	95.01
YOLACT ResNet50				98.94	90.90
YOLACT DarkNet53				96.46	96.46
YOLACT++ Resnet101	250	-	8	96.81	93.45
YOLACT++ Resnet101				96.84	94.64
YOLACT++ Resnet101	500	-	8	97.84	94.39
YOLACT++Resnet101				95.64	92.27
YOLACT++ Resnet101	250	0.001	8	97.82	90.97
YOLACT++Resnet101				95.80	93.76
YOLACT++ Resnet101	500	0.001	8	96.77	95.82
YOLACT++ Resnet101				95.79	92.55
YOLACT++ Resnet101	250	0.0001	8	96.52	94.41
YOLACT++ Resnet101				93.14	90.49
YOLACT++ Resnet101	500	0.0001	8	96.74	91.36
YOLACT++ Resnet101				95.48	92.93
YOLOv7-seg	250	0.0001	8	97.70	98.40
YOLOv8n-seg	250	0.0001	8	99.40	98.50
YOLOv8s-seg	250	0.0001	8	99.40	98.00
YOLOv8m-seg	250	0.0001	8	99.50	98.40

To ensure our segmentation model can provide the best performance in real time condition with US video, we combine such approach with two stages instance segmentation named Mask-RCNN. It can observe that our model outperformed all the state-of-the art as shown in Table 2. Based on such learning process, our proposed YOLOv8l-seg model produce the 99.50% mAP for Bbox and 98.40% mAP for mask higher than YOLOv7-seg, YOLACT and Mask-RCNN respectively as shown in Table 2. However, the mAP performance decrease when it tested in unseen image, 55.20% for bbox prediction and 34.70% for mask prediction as shown in Table 3. In addition, YOLOv8l-seg produce 68.4 fps for validation and 58.8 fps for testing faster than other method. Due to, YOLOv8l-seg has a more streamlined and efficient network architecture compared to previous versions (such as YOLOv7-seg). This allows YOLOv8l-seg to process images more quickly without sacrificing accurate. Due to, the mAP value of 55% when tested in unseen data is concerning, as it indicates that the model may not be performing at an acceptable level in real-world clinical practice. To addressing potential limitations of the model and improving its ability to perform well on unseen data, the image enhancement on US video is performed.

Table 2. The best segmentation model performance using validation data

Backbone	Val mAP (%)		Unseen mAP (%)		Inference time
	Bbox	Mask	Bbox	Mask	
YOLOv7-seg	97.70	98.40	37.70	28.06	57.2 fps
YOLOv8l-seg	99.50	98.40	55.20	34.70	68.4 fps
YOLACT (ResNet101)	98.61	93.44	54.54	36.56	17.2 fps
Mask-RCNN	89.31		33.74		-

Table 3. The mAP prediction after image enhancement

Backbone	mAP (%)		Inference time
	Bbox	Mask	
YOLOv7-seg	38.77	29.20	41.1 fps
YOLOv8l-seg	65.93	57.66	54.8 fps
YOLACT (ResNet101)	57.10	44.90	18.3 fps
Mask-RCNN	49.10		-

Both the bounding box loss and mask loss are used to train the YOLO algorithm to accurately detect and localize objects in images. By minimizing these losses, such should learn to predict accurate bounding boxes and masks, leading to better object detection performance. We have compared four losses of YOLOv7-seg, YOLOv8l-seg, YOLACT (one stages instance segmentation), and Maks-RCNN (two stages instance segmentation) to ensure the best segmentation performance as shown in Figure 3. The lower of the loss value, the better the model is performing, and the closer the predicted bounding boxes and masks are to the actual objects in the image. It can be observed that YOLOv8l-seg generate small loss without overfitting out performed other segmentation method. By observing the trend of the loss value over time that YOLOv8l-seg produce higher mAP and small loss, it means how well the such model in learning to segment and detect the fetal heart object.

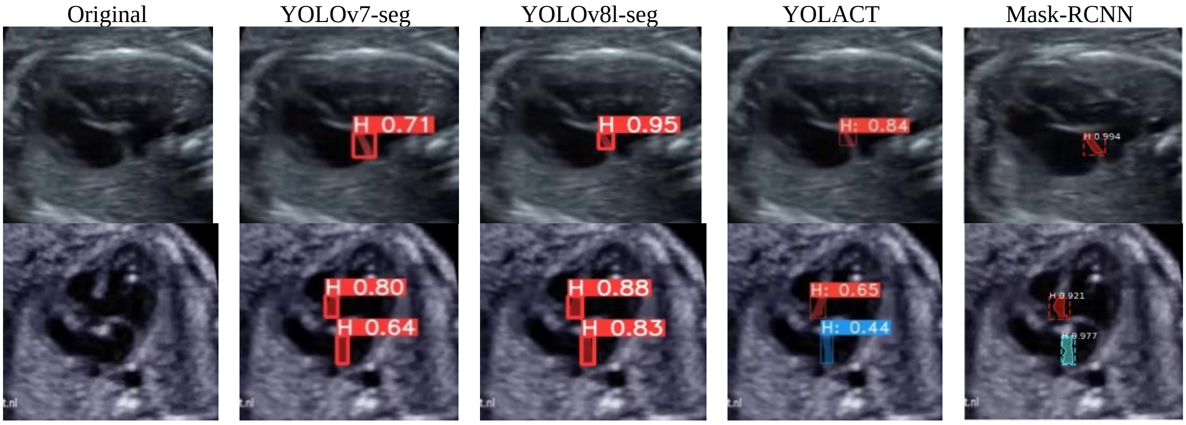
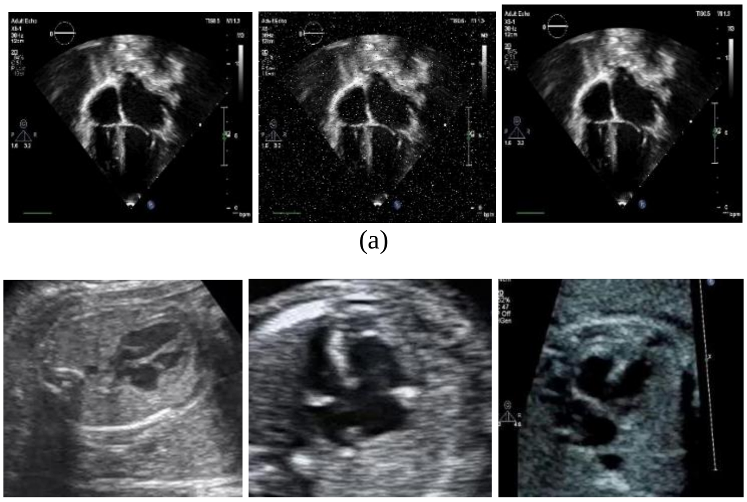


Figure 3. The segmentation prediction result based on the validation data for defect prediction in the atrial, ventricular and combine areas

MIRNet is designed to capture both low-level and high-level features of images, which allows it to effectively remove noise, sharpen edges, and enhance fine details in US images. It works by taking in a low-quality US image as input, and using its trained network to generate a high-quality output image. We trained our model using infant heart images due to their high quality in US imaging. To simulate real-world scenarios and make the model more robust, we intentionally decreased the quality of the infant heart images by injecting four types of noise, namely Gaussian, poisson, salt and pepper, and speckle. The image enhancement model was then generated using the MiRNet algorithm, as shown in Figure 4(a) and the prediction result in Figure 4(b).



(a)

(b)

Figure 4. The image enhancement process, (a) model generating with infant US video and (b) prediction result

SSIM takes into account both the local and global similarities between the images, which makes it suitable for image enhancement tasks such as denoising, deblurring, and super-resolution. By maximizing the SSIM score between the original and enhanced images, we can ensure that the enhanced image is visually similar to the original image and retains the important features while removing any noise or artifacts. As result found that the MIRNet model produced a SSIM of about 0.9824, and a PSNR varied with the amount of noise injected into the image. However, the overall PSNR achieved was over 33 dB, indicating that MIRNet was able to improve image quality with satisfactory performance. It can be seen in Table 3, after image enhancement the segmentation performance increase significant with past inference time. Testing with unseen data was difficult, as it could lead to a decrease in model performance. It can be observe that the use of MIRNet enhancement resulted in a decrease of approximately 33.57% in the predicted bbox and about 40.74% in the predicted mask. This led to a better predicted value as compared to the mAP value obtained before image enhancement. MIRNet has shown promising results in enhancing the quality of USimages and videos by reducing noise, improving contrast, and enhancing important image features. Moreover, MIRNet has the ability to learn and adapt to different types of noise, making it a powerful tool for US image and video enhancement. However, the performance of MIRNet may depend on factors such as the quality of the input data, the complexity of the enhancement task, and the specific implementation of the network.

4. DISCUSSION

In YOLO architecture, each object detected in an image is associated with a confidence value. This value represents the algorithm’s prediction about how certain it is that the detected object is actually present in the image. The confidence value is a score between 0 and 1, where a higher score indicates a higher level of confidence in the prediction. In this study, we set the confidence value is above a certain threshold over 0.5. The confidence value can be helpful in assessing the reliability of the object detection results. It can be seen in Figure 3, the confidence value of YOLOv8l-seg is higher than other, if the confidence value is very low, it may indicate that the algorithm is uncertain about the object detection and the result should be treated with caution. On the other hand, a high confidence value can provide greater confidence that the object detection is accurate.

The comparison of the model’s performance for predicting fetal heart defects is shown in Figure 5. We compare the performance of YOLOv7-seg, YOLOv8l-seg, and YOLACT before and after enhancement. Our study demonstrated that MIRNet is an effective DL architecture for enhancing the quality of US images, resulting in improved segmentation performance in a real-time. Our results suggest that MIRNet can effectively reduce noise and enhance important image features, leading to more accurate segmentation results. It can be observed that YOLOv8l-seg with MIRNet can improve the mAP from 55.22% to 65.93% for bbox prediction and from 34.70% to 57.66% for Mask prediction. It means, the mAP of bbox is increased about 14.73% and mAP mask labes is increased about 22.96% after image enhancement process as shown in Table 3. These findings highlight the potential of MIRNet as a valuable tool for improving the quality and reliability of US imaging in clinical settings.

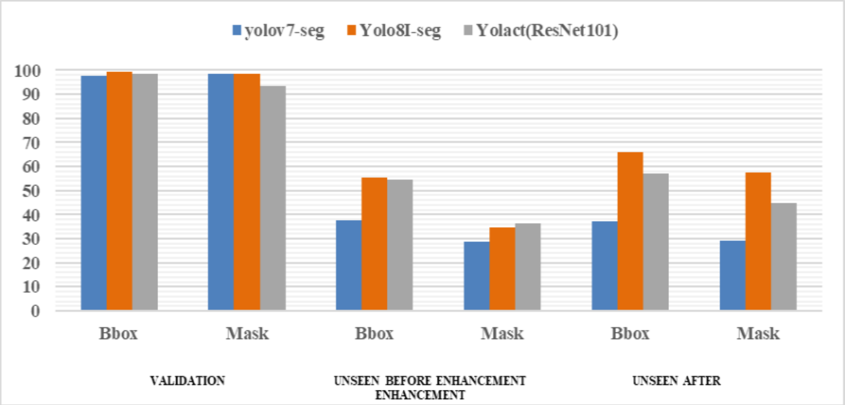


Figure 5. Model performance for fetal heart defect prediction with validation data, unseen data before and after MIRNet enhancement

This study performed a comparative analysis of YOLOv8l-seg with state-of-the-art models. However, real-time segmentation with the YOLOv8l-seg method is still limited, especially in medical imaging. To ensure a fair comparison, benchmarking was performed by evaluating the same method. We compared our results with three baseline YOLOv8l-seg models using the GRAZPEDWRI-DX dataset for pediatric wrist trauma fracture detection [25], the Nvidia AI City challenge for helmet detection [26], and custom data set for automated KI-67 proliferation and tumor-infiltrated lymphocyte estimation [27] as shown in Table 4. Compared to other methods, our approach yields better results. Unfortunetlly, our model runs at a much slower frame rate than the research conducted by Aboah *et al.* [26], nonetheless, our model produces real-time detection with 65.9% mAP, which is better than their method. The YOLOv8l-seg model is not only fast, accurate, and user-friendly, but it is also versatile and can be applied to a wide range of object detection and image segmentation tasks. However, its exceptional performance in medical imaging is particularly noteworthy because medical imaging often yields low-quality images. As a result, the YOLOv8l-seg model offers promising performance in medical object detection and segmentation.

Table 4. Our benchmarking of the YOLOv8l-seg model with MIRNet against the state-of-the-art

Author	Method	Dataset	mAP validation	mAP testing	fps
Aboah *et al.* [26]	Helmet detection	Nvidia AI City challenge dataset	93.0	58.6	95
Lapp *et al.* [27]	Automated KI-67 proliferation and tumor-infiltrated lymphocyte estim	Custom data set	55.3	-	-
Ju and Cai [28]	Fracture detection	GRAZPEDWRI-DX dataset Paediatric with wrist trauma	94.5	-	67.4
Proposed	Fetal heart defect detection	Custom dataset only fetal US video	99.5	55.20	68.4
		Custom dataset fetal US video with image enhancement	99.5	65.9	68.4

In the context of object detection algorithms, detecting small objects like fetal heart defects can be a challenge. This is because conventional object detection algorithms often rely on region proposals or sliding windows to identify objects, which can be computationally expensive and may miss small objects or objects with low contrast. However, there are one technique that can help improve the detection of small objects, using multi-scale detection with incorporating contextual information. While using our propose method has several advantages, there are still some limitations to consider. The limited availability of annotated data and poor image quality of fetal echocardiography images can also affect the performance in detecting fetal heart defects. Moreover, overfitting is a common problem in DL algorithms, and if YOLO is trained on a limited dataset, it may fail to generalize well to new images, leading to poor detection accuracy. Therefore, while our proposed model has demonstrated promising results for fetal heart defect detection using US video, we should continue to explore techniques to improve its performance. The impact on developing countries can be even greater, as there is a shortage of skilled operators in these regions, resulting in many women not receiving any US exams during their pregnancy [29]. Developing a system that can reduce the level of expertise required for scanning could have a profound impact. With such a system, individuals in remote areas with a basic anatomical background would be able to perform US exams. Only images with clinical value would be sent to radiologist experts for evaluation, regardless of the physician's location.

5. CONCLUSION

US is a widely utilized imaging modality globally for conducting first-line medical examinations during pregnancy. However, it is diagnostic performance still poses challenges due to the inherent characteristics of US imaging. Our study involves implementing the YOLOv8l-seg with MIRNet model for US image enhancement. The proposed model is a fully-convolutional architecture that utilizes a novel approach to feature learning. By combining contextual information from multiple scales, the model learns an enriched set of features that can enhance low-light images while preserving high-resolution spatial details. This approach is achieved through the use of advanced DL techniques, including real time instance segmentation and multi scale residual learning for image enhancement. The model facilitates information exchange between parallel streams, allowing high-resolution features to be consolidated with the help of low-resolution features, and vice versa. Overall, the YOLOv8l-seg with MIRNet model represents a powerful solution for enhancing low quality US images leading to better visual quality for achieving accurate diagnostic.

ACKNOWLEDGEMENTS

This work was supported by the Intelligent System Research Group (ISysRG) and was funded by the Ministry of Education, Culture, Research, and Technology as a Matching Fund, Kedai Reka Project. This work was also supported by Sriwijaya University funded by the Profesional Research 2023.

REFERENCES

[1] P. Quaresima *et al.*, "How to do a fetal cardiac scan," *Archives of Gynecology and Obstetrics*, vol. 307, no. 4, pp. 1269–1276, 2023, doi: 10.1007/s00404-023-06951-8.

[2] R. Tenajas, D. Miraut, C. I. Illana, R. A.-Gonzalez, F. A.-Valcayo, and J. L. Herraiz, "Recent advances in artificial intelligence-assisted ultrasound scanning," *Applied Sciences*, vol. 13, no. 6, 2023, doi: 10.3390/app13063693.

[3] P. G.-Canadilla, S. S.-Martinez, F. Crispi, and B. Bijnens, "Machine learning in fetal cardiology: what to expect," *Fetal Diagnosis and Therapy*, vol. 47, no. 5, pp. 363–372, 2020, doi: 10.1159/000505021.

[4] C. Maxwell and P. Glanc, "Imaging and obesity: a perspective during pregnancy," *American Journal of Roentgenology*, vol. 196, no. 2, pp. 311–319, 2011, doi: 10.2214/AJR.10.5849.

[5] J. Y. Mei and C. S. Han, "Ultrasound for the pregnant person with diabesity," *Clinical Obstetrics and Gynecology*, vol. 64, no. 1, pp. 144–158, 2021, doi: 10.1097/GRF.0000000000000600.

[6] L. Rundo *et al.*, "MedGA: A novel evolutionary method for image enhancement in medical imaging systems," *Expert Systems with Applications*, vol. 119, pp. 387–399, Apr. 2019, doi: 10.1016/j.eswa.2018.11.013.

[7] S. Agrawal, R. Panda, P. K. Mishro, and A. Abraham, "A novel joint histogram equalization based image contrast enhancement," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 4, pp. 1172–1182, Apr. 2022, doi: 10.1016/j.jksuci.2019.05.010.

[8] P. Gómez, M. Semmler, A. Schützenberger, C. Bohr, and M. Döllinger, "Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network," *Medical and Biological Engineering and Computing*, vol. 57, no. 7, pp. 1451–1463, 2019, doi: 10.1007/s11517-019-01965-4.

[9] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "MSR-net: low-light image enhancement using deep convolutional network," *arXiv-Computer Science*, pp. 1-9, 2017, doi: 10.48550/arXiv.1711.02488.

[10] Y. S. Chen, Y. C. Wang, M. H. Kao, and Y. Y. Chuang, "Deep photo enhancer: unpaired learning for image enhancement from photographs with GANs," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 2018, pp. 6306–6314. doi: 10.1109/CVPR.2018.00660.

[11] Y. Deng, C. C. Loy, and X. Tang, "Aesthetic-driven image enhancement by adversarial learning," in *MM 2018-Proceedings of the 2018 ACM Multimedia Conference*, ACM, 2018, pp. 870–878. doi: 10.1145/3240508.3240531.

[12] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 833–851. doi: 10.1007/978-3-030-01234-2_49.

[13] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: a deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017, doi: 10.1016/j.patcog.2016.06.008.

[14] W. Ren *et al.*, "Low-light image enhancement via a deep hybrid network," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4364–4375, Sep. 2019, doi: 10.1109/TIP.2019.2910412.

[15] L. H. Lee *et al.*, "Machine learning for accurate estimation of fetal gestational age based on ultrasound images," *NPJ Digital*

Medicine, vol. 6, no. 1, 2023, doi: 10.1038/s41746-023-00774-2.
[16] M. Komatsu *et al.*, "Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning," *Applied Sciences*, vol. 11, no. 1, pp. 1–12, 2021, doi: 10.3390/app11010371.
[17] S. W. Zamir *et al.*, "Learning enriched features for real image restoration and enhancement," in *Proceedings of the European conference on computer vision (ECCV)*, 2020, pp. 492–511. doi: 10.1007/978-3-030-58595-2_30.
[18] M. C. Fiorentino, F. P. Villani, M. Di Cosmo, E. Frontoni, and S. Moccia, "A review on deep-learning algorithms for fetal ultrasound-image analysis," *Medical Image Analysis*, vol. 83, 2023, doi: 10.1016/j.media.2022.102629.
[19] C. Lee *et al.*, "Development of a machine learning model for sonographic assessment of gestational age," *JAMA Network Open*, vol. 6, no. 1, 2023, doi: 10.1001/jamanetworkopen.2022.48685.
[20] B. S. Prabakaran, P. Hamelmann, E. Ostrowski, and M. Shafique, "FPUS23: an ultrasound fetus phantom dataset with deep neural network evaluations for fetus orientations, fetal planes, and anatomical features," *IEEE Access*, vol. 11, pp. 58308–58317, 2023, doi: 10.1109/ACCESS.2023.3284315.
[21] S. Nurmaini *et al.*, "Accurate detection of septal defects with fetal ultrasonography images using deep learning-based multiclass instance segmentation," *IEEE Access*, vol. 8, pp. 196160–196174, 2020, doi: 10.1109/ACCESS.2020.3034367.
[22] S. Nurmaini *et al.*, "Deep learning-based computer-aided fetal echocardiography: application to heart standard view segmentation for congenital heart defects detection," *Sensors*, vol. 21, no. 23, Nov. 2021, doi: 10.3390/s21238007.
[23] J. Terven, "A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023, doi: 10.3390/make5040083.
[24] A. I. Sapitri *et al.*, "Deep learning-based real time detection for cardiac objects with fetal ultrasound video," *Informatics in Medicine Unlocked*, vol. 36, 2023, doi: 10.1016/j.imu.2022.101150.
[25] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-time instance segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, 2019, pp. 9156–9165. doi: 10.1109/ICCV.2019.00925.
[26] A. Aboah, B. Wang, U. Bagci, and Y. A.-Gyamfi, "Real-time multi-class helmet violation detection using few-shot data sampling technique and YOLOv8," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, 2023, pp. 5350–5358. doi: 10.1109/CVPRW59228.2023.00564.
[27] S. Z. Lapp, E. David, and N. S. Netanyahu, "PathRTM: Real-time prediction of KI-67 and tumor-infiltrated lymphocytes," *Arxiv-Quantitative Biology*, pp. 1-12, 2023, doi: 10.48550/arXiv.2305.00223.
[28] R. Y. Ju and W. Cai, "Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm," *Scientific Reports*, vol. 13, no. 1, Nov. 2023, doi: 10.1038/s41598-023-47460-7.
[29] S. Shah, B. A. Bellows, A. A. Adedipe, J. E. Totten, B. H. Backlund, and D. Sajed, "Perceived barriers in the use of ultrasound in developing countries," *Critical Ultrasound Journal*, vol. 7, no. 1, 2015, doi: 10.1186/s13089-015-0028-2.

BIOGRAPHIES OF AUTHORS

Sutarno is currently a lecturer and a researcher with the Intelligent System Research Group, Faculty of Computer Science, Sriwijaya University, Indonesia. His research interests include medical imaging, biomedical signal processing, deep learning, and machine learning. He can be contacted at email: sutarno@unsri.ac.id.

Siti Nurmaini (member, IEEE) received the master's degree in control system from the Bandung Institute of Technology (ITB), Indonesia, in 1998, and the Ph.D. degree in computer science from the University of Technology Malaysia (UTM), in 2011. She is currently a professor with the Faculty of Computer Science, Universitas Sriwijaya. Her research interests include biomedical engineering, deep learning, machine learning, image processing, control systems, and robotic. She can be contacted at email: sitinurmaini@gmail.com or siti_nurmaini@unsri.ac.id.

Ade Iriani Sapitri is currently a doctoral student at the Faculty of Engineering, Sriwijaya University, Indonesia. In addition to her studies, she is also a Research Assistant affiliated with the Intelligent System Research Group. Her research interests include medical imaging, deep learning, and machine learning. She can be contacted at email: adeirianisapitri13@gmail.com.

Muhammad Naufal Rachmatullah is currently a lecturer and a researcher with the Intelligent System Research Group, Faculty of Computer Science, Sriwijaya University, Indonesia. His research interests include medical imaging, biomedical signal processing, deep learning, and machine learning. He can be contacted at email: naufalrachmatullah@gmail.com.

Bambang Tutuko received the master’s degree in control system from the Bandung Institute of Technology (ITB), Indonesia, in 1998, and the Doctor degree in computer science from the Sriwijaya University, in 2019. He is currently a Professor with the Faculty of Computer Science, Sriwijaya University. His research interests include biomedical engineering, deep learning, machine learning, image processing, control systems, and robotic. He can be contacted at email: bambangtutuko60@gmail.com.

Annisa Darmawahyuni is currently a lecturer and a researcher with the Intelligent System Research Group, Faculty of Computer Science, Sriwijaya University, Indonesia. Her research interests include biomedical signal processing, deep learning, and machine learning. She can be contacted at email: riset.annisadarmawahyuni@gmail.com.

Firdaus is currently a lecturer and a researcher with the Intelligent System Research Group, Faculty of Computer Science, Sriwijaya University, Indonesia. His research interests include pattern classification, text analysis, bibliographic system, cardiovascular system, classification, computer aided instruction, control engineering computing, data mining, diseases, natual language processing. He can be contacted at email: virdauz@gmail.com.

Anggun Islami is currently a lecturer and a researcher with the Intelligent System Research Group, Faculty of Computer Science, Sriwijaya University, Indonesia. Her research interests include pattern classification, text analysis, bibliographic system, cardiovascular system, classification, data mining. She can be contacted at email: anggunislami2@gmail.com.

Samsuryadi is currently a lecturer and a researcher Department of Informatic Engineering, Faculty of Computer Science, Sriwijaya University. His research interests include medical imaging, biomedical signal processing, deep learning, and machine learning. He can be contacted at email: syamsuryadi@unsri.ac.id.