

Enhancing ultrasound image quality using deep structure of residual network

Ade Iriani Sapitri, Siti Nurmaini, Muhammad Naufal Rachmatullah, Annisa Darmawahyuni,
Firdaus Firdaus, Anggun Islami, Bambang Tutuko, Akhiar Wista Arum

Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia

Article Info

Article history:

Received Sep 2, 2024

Revised Apr 5, 2025

Accepted May 24, 2025

Keywords:

Deep learning

Enhancement

Medical image

Residual network

Ultrasonography

ABSTRACT

Ultrasonography, a medical imaging technique, is often affected by various types of noise and low brightness, which can result in low image quality. These drawbacks can significantly impede accurate interpretation and hinder effective medical diagnoses. Therefore, improving image quality is an essential aspect of the field of ultrasound systems. This study aims to enhance the quality of ultrasound images using deep learning (DL). The experiment is conducted using a custom dataset consisting of 2,175 infant heart ultrasound images collected from Indonesian hospitals, and the model is subsequently generalized using other datasets. We propose enhanced deep residual network combined convolutional neural networks (EDR-CNNs) to improve the image quality. After the enhancement process, our model achieved peak signal-to-noise ratio (PSNR) and structural similarity index metrics (SSIM) scores of 38.35 and 0.92 respectively, outperforming other methods. The benchmarking with other ultrasound medical images indicates that our proposed model produces good performance, as evidenced by higher PSNR, lower SSIM, a decrease in mean square error (MSE), and a lower contrast improvement index (CII). In conclusion, this study encapsulates the forthcoming trends in advancing low-illumination image enhancement, along with exploring the prevailing challenges and potential directions for further research.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Siti Nurmaini

Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya

Palembang, South Sumatera, Indonesia

Email: siti_nurmaini@unsri.ac.id

1. INTRODUCTION

Enhancing the quality of medical images, particularly in ultrasound systems, is a crucial area of focus in the medical field. This is primarily attributed to the extensive utilization of ultrasound technology in medical diagnosis and a wide range of medical procedures. Ultrasound images often suffer from issues such as noise and blur, which can significantly impede medical interpretation and accurate diagnosis [1]–[4]. Several factors contribute to the low quality of ultrasound images. One of these factors is acoustic variability, where the ultrasound image is affected by acoustic variations in the body, causing noise in the image [5]–[8]. In addition, signal deficiencies or attenuations may occur as the ultrasound signal passes through different body tissues, and resolution limitations may arise due to patient movement, suboptimal positioning, or the use of less sophisticated ultrasound technology [9]–[11]. These problems can significantly affect the quality of ultrasound images and make medical interpretation and diagnosis challenging. To address this problem, many research studies have focused on developing methods to improve the quality of ultrasound images.

There are two broad categories of traditional methods for low-light enhancement such as histogram equalization [11], [12] and Retinex models [13], [14]. The histogram technique aims to enhance the contrast of an image by redistributing the intensity values of the pixels, resulting in a brighter and more vibrant image. Whereas a typical approach based on Retinex models decomposes a low-light image into two components: the reflection component and the illumination component using priors or regularizations. The estimated reflection component is considered the enhanced result. However such two methods have several limitations [14], i) Both histogram equalization and Retinex models can be particularly problematic when dealing with images containing sharp edges and textures, as they may become exaggerated and appear artificial; ii) In such approaches, the existing noise in the images is often disregarded, leading to either its persistence or amplification in the enhanced results; iii) Complex optimization processes involved in Retinex models and the computation of image histograms and cumulative distribution functions in histogram equalization increasing the computational time, and iv) Both approaches necessitate manual adjustments and interventions to accommodate specific scenarios or edge cases.

In recent times, profound advancements have been made in the realm of image enhancement techniques, specifically leveraging DL techniques such as convolutional neural networks (CNNs) [15], [16]. These remarkable developments have shown great potential in reducing noise, augmenting resolution, and mitigating artifacts. As a result, CNNs have emerged as highly promising solutions to tackle the persistent challenge of enhancing ultrasound image quality, ultimately leading to a substantial improvement in diagnostic accuracy. There has been research focused on developing more specialized and effective DL techniques to improve ultrasound image quality, such as using transfer learning on existing DL architectures [12]–[14], [17]–[20]. These advances continue to improve the quality of ultrasound images and improve the accuracy of medical diagnoses and interventions. This research contributes as follows:

- Proposing a custom deep residual CNNs architecture to improve the quality of ultrasound images in terms of both performance and efficiency.
- Validating the effectiveness of our proposed model using rigorous experiments and benchmarks against established state-of-the-art DL models.
- Demonstrating the proposed custom architecture with four ultrasound datasets i.e., infant heart, fetal heart, fetal head, and abdomen.
- Implementing the proposed model for predicting the fetal heart substructure abnormality.

2. MATERIALS AND METHOD

The general methodology of our experiment is depicted in Figure 1, consisting of four main stages. The first stage, data acquisition is carried out. Data for the acquisition was obtained from several previously collected ultrasound videos. The data is then extracted based on the frame and cropped the heart object directly. The second stage involves data pre-processing, where low-resolution data using noise injection and high resolution are generated as targets. This process involves labeling the data for further analysis. In the third stage, the proposed deep learning algorithm proposes an enhanced residual CNN architecture. The final stage involves performance measurement, where the peak signal-to-noise ratio (PSNR), structural similarity index metrics (SSIM), mean square error (MSE), and contrast improvement index (CII) values are calculated to evaluate the performance of the proposed deep-learning model. The whole steps constitute the general methodology of our experiment.

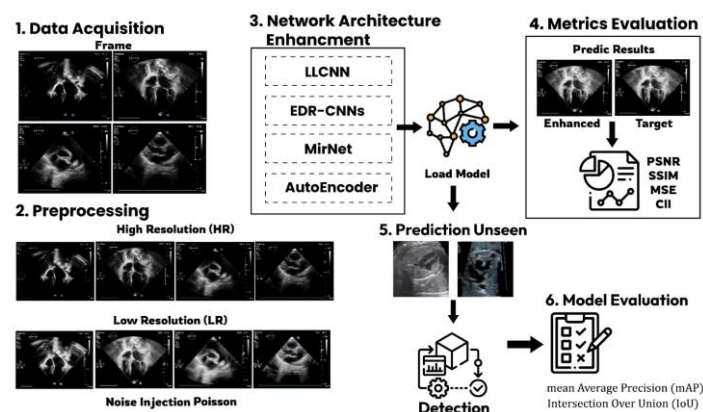


Figure 1. The general methodology of our experiment

2.1. Data acquisition

The dataset used in this research consists of medical images acquired from ultra-sound devices, including images of infant hearts, fetal hearts, fetal heads, and abdomens. While infant and fetal heart datasets are custom-made from Indonesian hospitals, the remaining datasets are sourced from public repositories. The data distribution of our proposed model can be found in Table 1. The study was carried out in two main stages: i) In the first stage, we focused on generating an image enhancement model using images of infant hearts, and ii) Subsequently, the enhancement model that demonstrated the best performance was selected and applied to the other datasets for benchmarking purposes.

Table 1. Data distribution for the learning process.

USG Object	Training	Validation	Testing	Unseen	Total
Infant's heart	1177	732	244	60	2213
Fetal heart	410	78	26	28	542
Fetal head	915	279	84	56	1334
Abdomen	472	132	84	23	711

The custom of the infant heart dataset is obtained from a general hospital, Dr. Mohammad Hoesin, Indonesia. Videos are captured using a Philips EPIQ Elite ultrasound machine and video file sizes range from 30-50 MB. To generate a reliable image enhancement model, we used 1,171 training data and 976 validation data. All the image data used to build the image enhancement model originated from images of a child's heart. It can be seen a sample image of an infant's heart used for model enhancement development in Figure 2.

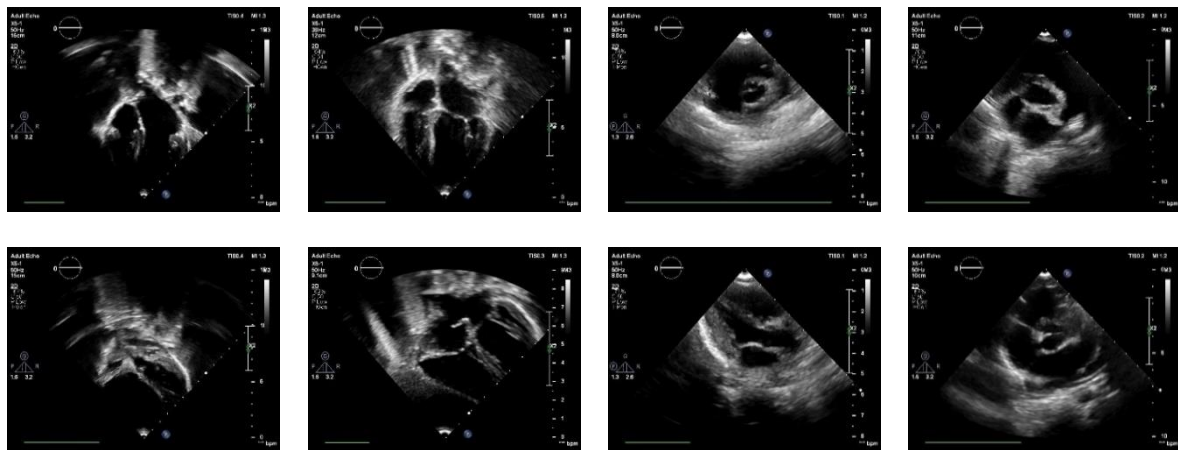


Figure 2. A sample image of an infant's heart as raw data for the development of the model enhancement

2.2. Data pre-processing

Based on the data acquisition process, previously collected several ultrasound videos. Data is extracted frame by frame, and all data is cropped to 800×600 size. Then, the dataset is processed by producing high-resolution and low-resolution images as input for model training as shown in Figure 3. High-resolution images are the target output from the image processing process and low-resolution images is generated as the input for the image enhancement model. Low-resolution images as input models use noise injection which often occurs in medical ultrasound data that can give the raw image a smoother or more blurry appearance. Different types of noise are added, such as poison, salt-and-pepper noise, or speckle noise. We control the intensity and type of noise to add based on the different settings to achieve the desired level of noise. High-resolution images serve as the target result of image processing, while low-resolution images are the input needed to train the model to learn patterns in low-resolution data and improve image quality. The results of the pre-processing stage are visualized in Figure 4. After low-resolution images are entered into the model, a multiscale information representation process is performed, combining information from various spatial scales and extracting relevant features to improve image quality. By taking advantage of low and high-resolution image types, the enhancement model can effectively improve image quality and produce high-quality images.

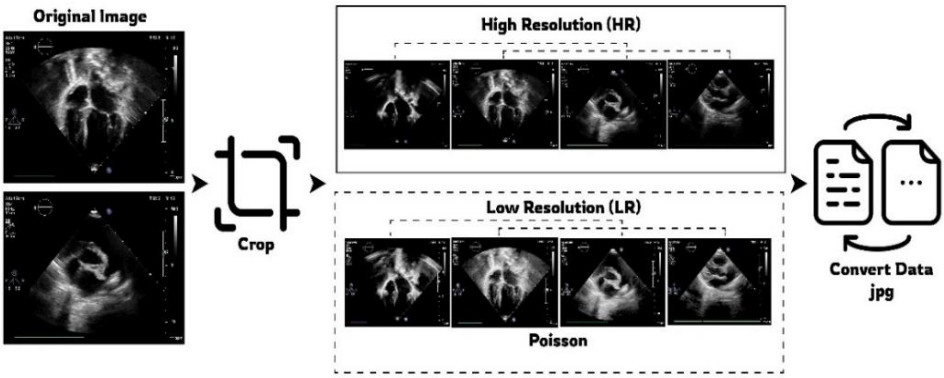


Figure 3. Data pre-processing process

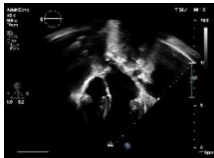
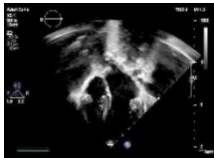
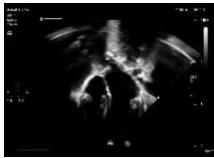
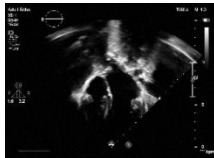
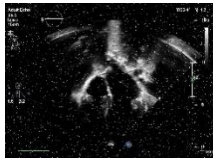
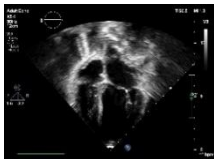

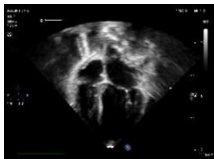
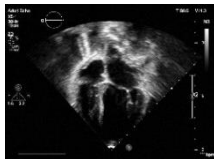
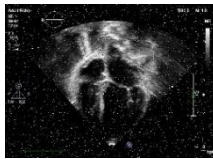



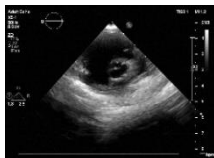






Original image (high resolution)	The low-resolution image by adding a controlled amount of noise			
	Poisson	Speckle	Gaussian	Salt Pepper
-	Lam = 3.0, size = (5, 5)	MedianBlur (img, 5)	Mean = 0, variance = 50, and sigma = variance * 0.5	Salt =1 (255, 255, 255) and Pepper = 2 (0,0,0)
				
				
				
				

Figure 4. High-resolution and low-resolution

2.3. Image enhancement model

This study proposed customized enhanced deep residual network combined convolutional neural networks (EDR-CNNs) architecture as shown in Figure 5. The structure is modified by emulating residual network (ResNet) which is designed to address the challenge of image noise [14], [15]. The EDR-CNN structure focuses on convolution widening, enabling networks to expand their visual range without compromising image resolution. It uses multi-resolution branches with different resolutions to handle information at various scales. The network bypass connections enable the learning of more complex features and accelerate network convergence.

In order to improve model performance and efficiency, the modified structure includes deeply separable attention and convolution modules. The network structure is shown in Table 2. Deep convolution is used to reduce the computational complexity in the convolutional layer and reduce the number of parameters

required by the model. In addition, the proposed model has been modified with a sampling technique called sub-pixel convolution to increase the image resolution after super-resolution processing. This technique estimates new pixels and studies existing spatial patterns in low-resolution image quality. The structure of the EDR-CNNs model comprises convolutions with 3×3 filters followed by the LeakyReLU activation function and Charbonnier Loss, as (1):

$$L(x, y) = ((x - y)^2 + \epsilon^2)^{\frac{p}{2}} \quad (1)$$

where x and y are the input and target images, respectively. ϵ is a small constant to avoid division by zero.

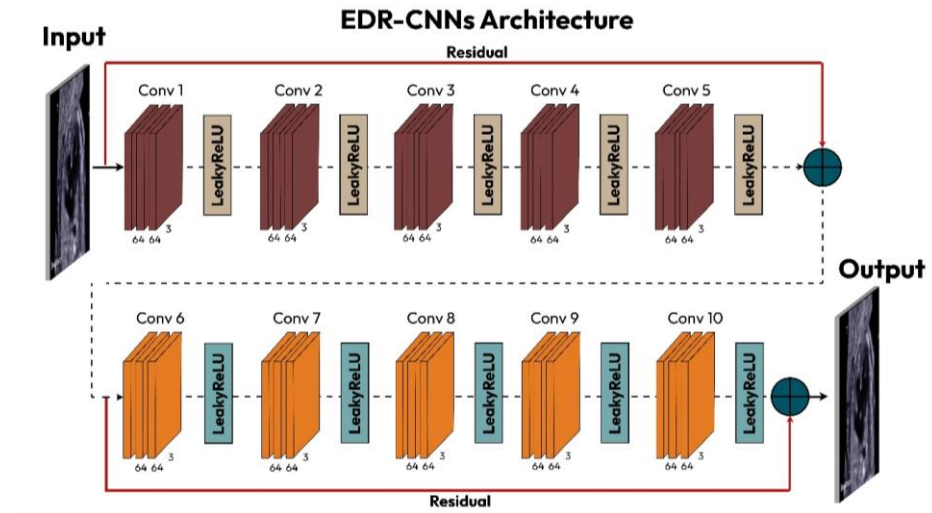


Figure 5. The proposed EDR-CNNs architecture

Table 2. The proposed EDR-CNNs network

Convolution	Layer (type)	Output Shape	Number of Parameters
Input Image	Input_1 (InputLayer)	[(None, None, None, 3)]	0
Conv 1	Conv2d (Conv2D)	[(None, None, None, 64)]	1792
	leaky_re_lu (LeakyReLU)	[(None, None, None, 64)]	0
Conv 2	conv2d_1 (Conv2D)	[(None, None, None, 64)]	36928
	leaky_re_lu_1 (LeakyReLU)	[(None, None, None, 64)]	0
Conv 3	conv2d_2 (Conv2D)	[(None, None, None, 64)]	36928
	leaky_re_lu_2 (LeakyReLU)	[(None, None, None, 64)]	0
Conv 4	conv2d_3 (Conv2D)	[(None, None, None, 64)]	36928
	leaky_re_lu_3 (LeakyReLU)	[(None, None, None, 64)]	0
Conv 5	conv2d_4 (Conv2D)	[(None, None, None, 64)]	36928
	leaky_re_lu_4 (LeakyReLU)	[(None, None, None, 64)]	0
Residual	add_1 (Add)	[(None, None, None, 3)]	0
Conv 6	conv2d_5 (Conv2D)	[(None, None, None, 64)]	36928
	leaky_re_lu_5 (LeakyReLU)	[(None, None, None, 64)]	0
Conv 7	conv2d_6 (Conv2D)	[(None, None, None, 64)]	36928
	leaky_re_lu_6 (LeakyReLU)	[(None, None, None, 64)]	0
Conv 8	conv2d_7 (Conv2D)	[(None, None, None, 64)]	36928
	leaky_re_lu_7 (LeakyReLU)	[(None, None, None, 64)]	0
Conv 9	conv2d_8 (Conv2D)	[(None, None, None, 64)]	36928
	leaky_re_lu_8 (LeakyReLU)	[(None, None, None, 64)]	0
Conv 10	conv2d_9 (Conv2D)	[(None, None, None, 3)]	1731
Residual	add_1 (Add)	[(None, None, None, 3)]	0

After several convolution layers there is residual linking which is done by adding the output of the first layer to the output of the last layer before the next convolution layer. This architectural design helps in studying more detailed features and preserving the original information of the input. The final output goes through the last convolution layer and the sigmoid activation function to produce the final output which is the enhanced image prediction. The total number of parameters in the model is 263,971 which also includes the weights and biases in each convolution layer. The proposed model is generated on hardware equipped with TensorFlow 2.7 and NVIDIA RTX 2080ti GPU. All data has an initial input size of 128×128 for the training and validation process. The model was trained using Adam's loss function and optimizer with a learning rate of 0.0001. Model enhanced deep residual CNNs are based on multiple residual blocks using a batch normalization layer for consistent results. Our implementation includes only 16 remaining blocks with 64 channels.

2.4. Evaluation of image quality metrics

The evaluation step is used to compare the performance of the proposed method with the image quality improvement method. The peak signal-to-noise ratio (PSNR) is used to measure the quality of the reconstructed image. Mathematically [21], PSNR is expressed by (2),

$$PSNR = 10 \cdot \log_{10} \left(\frac{\max(I)^2}{MSE} \right) \quad (2)$$

where I represents each image and MSE is the mean square error. The lower the MSE, the higher the PSNR. SSIM is an image processing technique that measures the structural similarity between the actual image and the reconstructed image, such as image compression or image up-sampling/down-sampling [17]. Mathematically, SSIM can be represented by (3),

$$SSIM(I, \bar{I}) = \frac{(2\mu_I\mu_{\bar{I}} + c_1)(2\sigma_{I\bar{I}} + c_2)}{(\mu_I^2 + \mu_{\bar{I}}^2 + c_1)(\sigma_I^2 + \sigma_{\bar{I}}^2 + c_2)} \quad (3)$$

The equation (4) represents the per pixel comparison between the two images for SSIM. MSE has a duty as image recovery or image enhancement. MSE measures the average of the squared difference between the original pixel value and the predicted pixel value in the enhanced image.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (4)$$

The calculation process involves calculating the squared difference between the original pixels and the predicted pixels where n is the data point number Y_i are the original pixels of the image and \hat{Y}_i is the pixel prediction from the model. The CII evaluates the competitiveness of various contrast enhancement techniques, the most well-known benchmark being the image enhancement measure. It is used to compare the results of contrast enhancement methods. Contrast enhancement can be measured using CII as a ratio. CII is defined in (5) as:

$$CII = \frac{C_{proposed}}{C_{original}} \quad (5)$$

where C is the average value of the local contrast measured with a 3×3 window as:

$$\frac{\max - \min}{\max + \min} \quad (6)$$

$C_{proposed}$ dan $C_{original}$ is the average value of the local contrast in the output and original images, respectively.

3. RESULTS AND DISCUSSION

3.1. Model enhancement with infant heart object

In this section, experiments are conducted to create an image enhancement model that yields satisfactory performance in terms of PSNR, SSIM, MSE, and CII. To build such a model, only infant heart objects are used. Curating an infant heart dataset for constructing high-fidelity DL models tailored for clinical deployment necessitates images of sufficiently high quality. This enables the algorithm to discern both

common and distinct patterns among images and facilitates precise measurements. Five DL model structures have been developed, and we will assess the best performance of the selected model in Table 3. The entire image was trained by utilizing small 256×256 patches extracted from infant heart images. The training process spanned 50 epochs, employing a batch size of 4. Notably, the EDR-CNNs architecture was implemented, incorporating the Charbonnier loss as the loss function. For optimization, the Adam Optimizer was utilized, with a learning rate set at 0.0001.

Table 3. The five models generated to select the best structure

CNNs	ResNet	Model 1	Model 2	EDR-CNNs
16 layer	50 layer	12 layer	10 layer	10 layer
Input Image	Input Image	Input Image	Input Image	Input Image
Conv 1	Conv 1	Conv 1	Conv 1	Conv 1
Batch Normalization	Conv 2.x	Conv 2	Conv 2	LeakyReLU
ReLU	Batch Normalization	Conv 3	Conv 3	Conv 2
Conv n	ReLU	Conv 4	Conv 4	Conv 3
Batch Normalization	Addition (Residual)	ReLU	Conv 5	LeakyReLU
Addition (Residual)	Conv 3.x	Addition (Residual)	Addition (Residual)	Conv 4
ReLU	Batch Normalization	Conv 5	Conv 6	LeakyReLU
Conv n	ReLU	Conv 6	Conv 7	Conv 5
Batch Normalization	Addition (Residual)	Conv 7	Conv 8	LeakyReLU
Addition (Residual)	Conv 4.x	Conv 8	Conv 9	Addition (Residual)
ReLU	Batch Normalization	ReLU	Conv 10	Conv 6
Conv n	ReLU	Addition (Residual)	Addition (Residual)	LeakyReLU
Batch Normalization	Addition (Residual)	Conv 9	Activation Sigmoid	Conv 7
Addition (Residual)	Conv 5.x	Conv 10	Loss charbonnier	LeakyReLU
ReLU	Batch Normalization	Conv 11	Optimizer Adam	Conv 8
Conv n	ReLU	Conv 12	-	LeakyReLU
Batch Normalization	Addition (Residual)	ReLU	-	Conv 9
Addition (Residual)	Activation Softmax	Addition (Residual)	-	LeakyReLU
ReLU	Loss MSE	Activation ReLU	-	Conv 10
Activation ReLU	Optimizer Adam	Loss MSE	-	Residual
Loss MSE	-	Optimizer Adam	-	Activation Sigmoid
Optimizer Adam	-	-	-	Loss Charbonnier
-	-	-	-	Optimizer Adam

Through the analysis of the five model structures, we observe diverse performance outcomes in PSNR, SSIM, MSE, and CII values. As found in the result, the custom EDR-CNNs structure stands out by surpassing the performance of the other models. The PSNR values approach 40 dB, and it demonstrates comparable performance in other metrics as well in Figure 6. This compelling performance suggests that the proposed model warrants further evaluation to ascertain its resilience and ability to generalize across various medical objects.

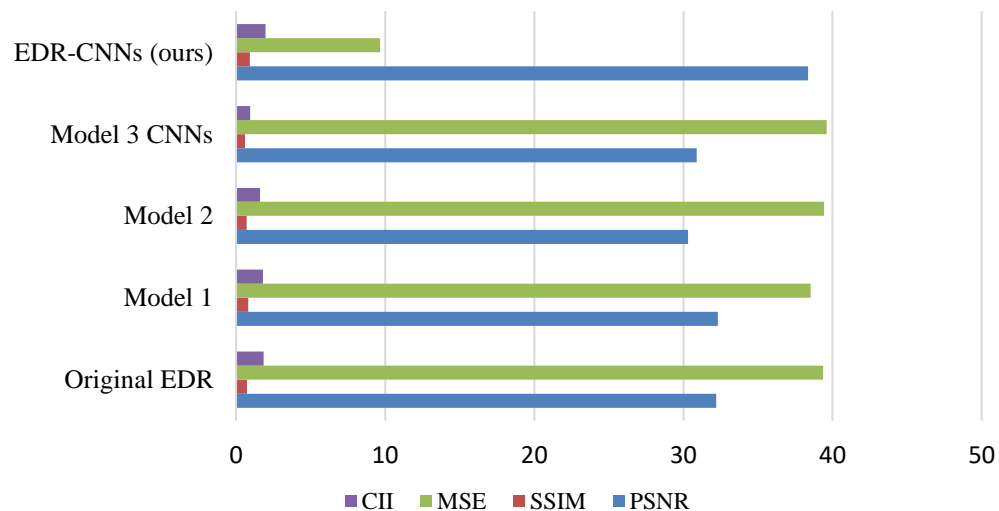


Figure 6. The performance validation of the selected model

3.2. Model evaluation

This section pertains to the experimental report that focuses on the evaluation and interpretation of our model's performance. We benchmark a thorough analysis of the EDR-CNNs obtained results against various methods, including contrast limited adaptive histogram equalization (CLAHE) [21], bilateral filter [22], [23], MirNet, low-light convolutional neural network (LLCNN) [24], and Autoencoder [25] as seen in Table 4. Four performance values were used to indicate the best image enhancement architecture. Higher PSNR values indicate better image quality, SSIM values range from -1 to 1, where 1 indicates identical images, lower MSE values indicate better image quality and higher CII values indicate better color image quality. We can observe that the EDR-CNNs produce outstanding results, with 38.35 PSNR, 0.92 SSIM, 9.66 MSE, and 1.97 CII. All the test values demonstrate satisfactory results in comparison to the other five methods.

Table 4. Model image enhancement performance with CLAHE, Bilateral filter, MirNet, LLCNN, Autoencoder, and EDSR-CNNs

Metrics	CLAHE	Bilateral Filter	MirNet	LLCNN	Autoencoder	EDR-CNNs
PSNR	31.92	37.76	35.81	37.78	36.10	38.35
SSIM	0.55	0.94	0.76	0.91	0.898	0.92
MSE	42.01	11.06	24.58	11.01	16.14	9.66
CII	1.58	0.99	0.98	1.36	0.94	1.97

The graphs for loss and PSNR during the training and validation process, indicate a consistent improvement in the quality of medical images throughout the enhancement process. Figure 7(a) to 7(d) visualizes the results of the model comparison using MirNet, LLCNN, Autoencoder, and EDR-CNNs. The whole graphical representation highlights a notable trend in the PSNR values, demonstrating an upward trajectory signifying enhanced image quality. It can be observed that the PSNR value varies with each iteration, but it should be noted that the PSNR value given to the EDR-CNNs method does not show significant fluctuations in Figure 7(d). As the model improves, the PSNR values should increase, indicating better image fidelity, as well as loss. The loss curve displays a gradual decline, eventually converging to zero with a smooth progression. A lower loss in EDR-CNNs response indicates that the model's predictions are getting closer to the actual values. It can be observed that the EDR-CNNs model effectively learns and adapts its parameters, leading to a desirable decrease in the loss.

Figure 8(a) to 8(g), showcased the heightened image quality achieved through the implementation of several image enhancement models. It can be seen from the figure that all low-illumination image enhancement methods based on deep learning effectively improve the accuracy of an infant's heart. But the visual effect is different. These techniques showcase the advancements in using deep learning to improve the quality of images by enhancing details, reducing noise, and mitigating other image-related issues. Each approach has its unique characteristics and applications within the realm of image enhancement and restoration. Nevertheless, when it comes to PSNR, SSIM, and MSE performance, EDR-CNNs surpass other approaches.

In this section, we undertook a comprehensive benchmarking analysis using the latest advancements in the field. This analysis involves a meticulous evaluation and comparison of our proposed model against well-established deep-learning approaches within the domain of enhancement. To facilitate this assessment, we employed four distinct datasets: the infant heart dataset for the development of our enhancement model, and three additional datasets — fetal heart, fetal head, and abdomen — to thoroughly evaluate the model's performance. To ensure the fairness of the test, all the methods are tested under the same hardware environment after reaching the optimal level of training. The experimental outcomes yielded noteworthy insights, specifically in terms of PSNR, SSIM, and MSE metrics across all four datasets in detail for various techniques refer to Figures 9(a) to 9(c). It observes that our proposed model EDR-CNNs consistently delivers satisfactory results across the entire dataset, achieving higher PSNR, lower SSIM, and decreased MSE values. This indicates that our suggested model consistently produces favorable performance. By evaluating the enhanced images using different techniques and datasets, it can be stated that the results are promising.

To ensure the performance of the proposed image enhancement technique using EDR-CNNs, such a model has been implemented for segmenting and detecting the 10 objects of the fetal heart substructure, including the left atrium (LA), right atrium (RA), left ventricle (LV), right ventricle (RV), tricuspid valve (TV), pulmonary valve (PV), mitral valve (MV), interventricular septum (IVS), aorta (Ao) and hole (defect). The experiments involved the utilization of the you only look at once with masking (YOLACT) framework. YOLACT is an algorithm designed for real-time object detection that combines object detection with instance segmentation. Its goal is to concurrently identify and segment objects present in an image. Detecting the fetal heart sub-structure object proves to be highly challenging due to its small size, the poor quality of

images, and its shape that dynamically changes due to fetal movements [5]. While it often performs well during testing using validation data, its effectiveness tends to decline or even result in failure when faced with unseen data [6], [7]. The image sample as visualized in Figure 10.

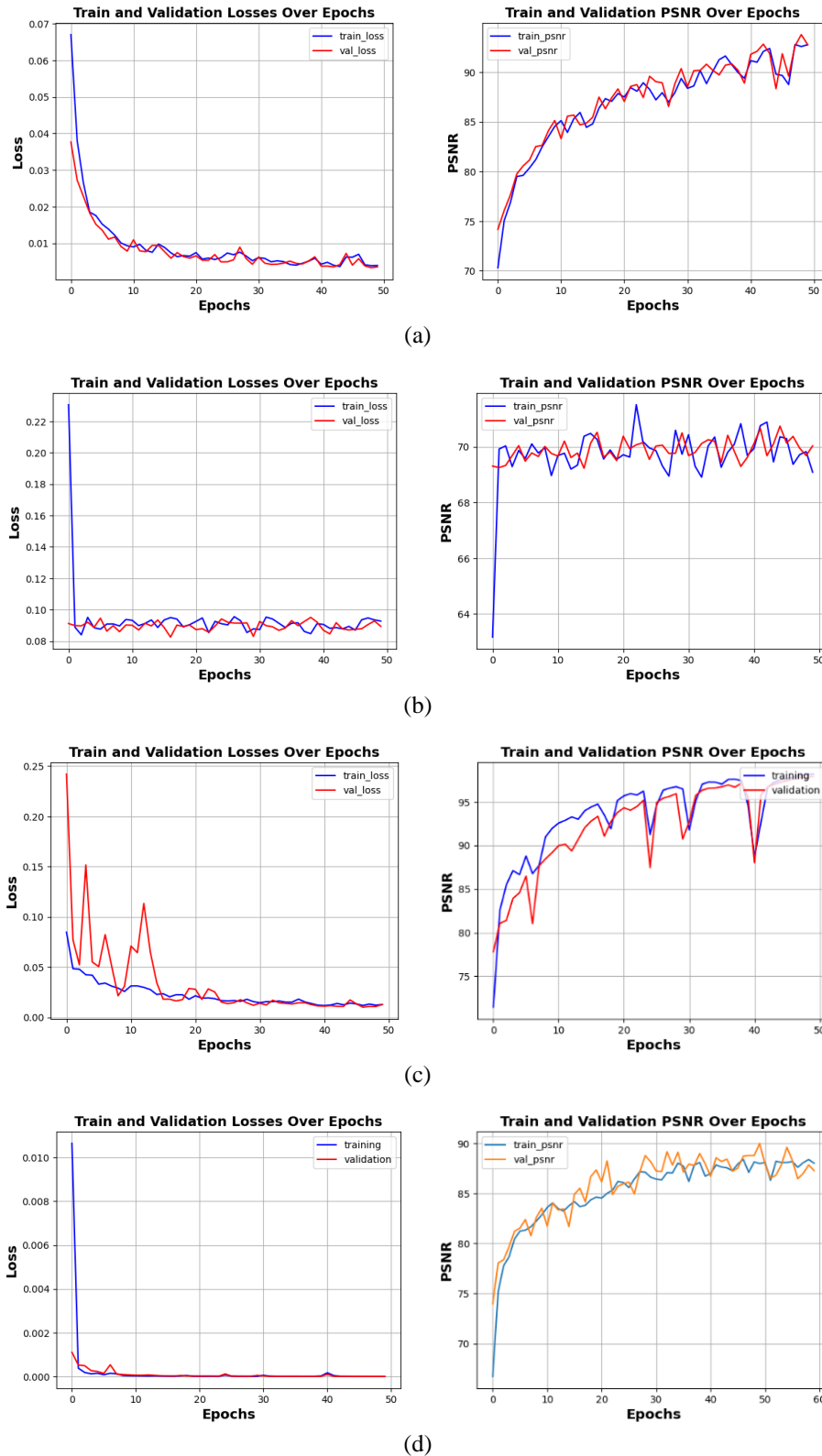


Figure 7. Graphic training and validation of loss and PSNR over epochs for infant heart object including (a) MirNet, (b) LLCNN, (c) Autoencoder, and (d) EDR-CNNs structure

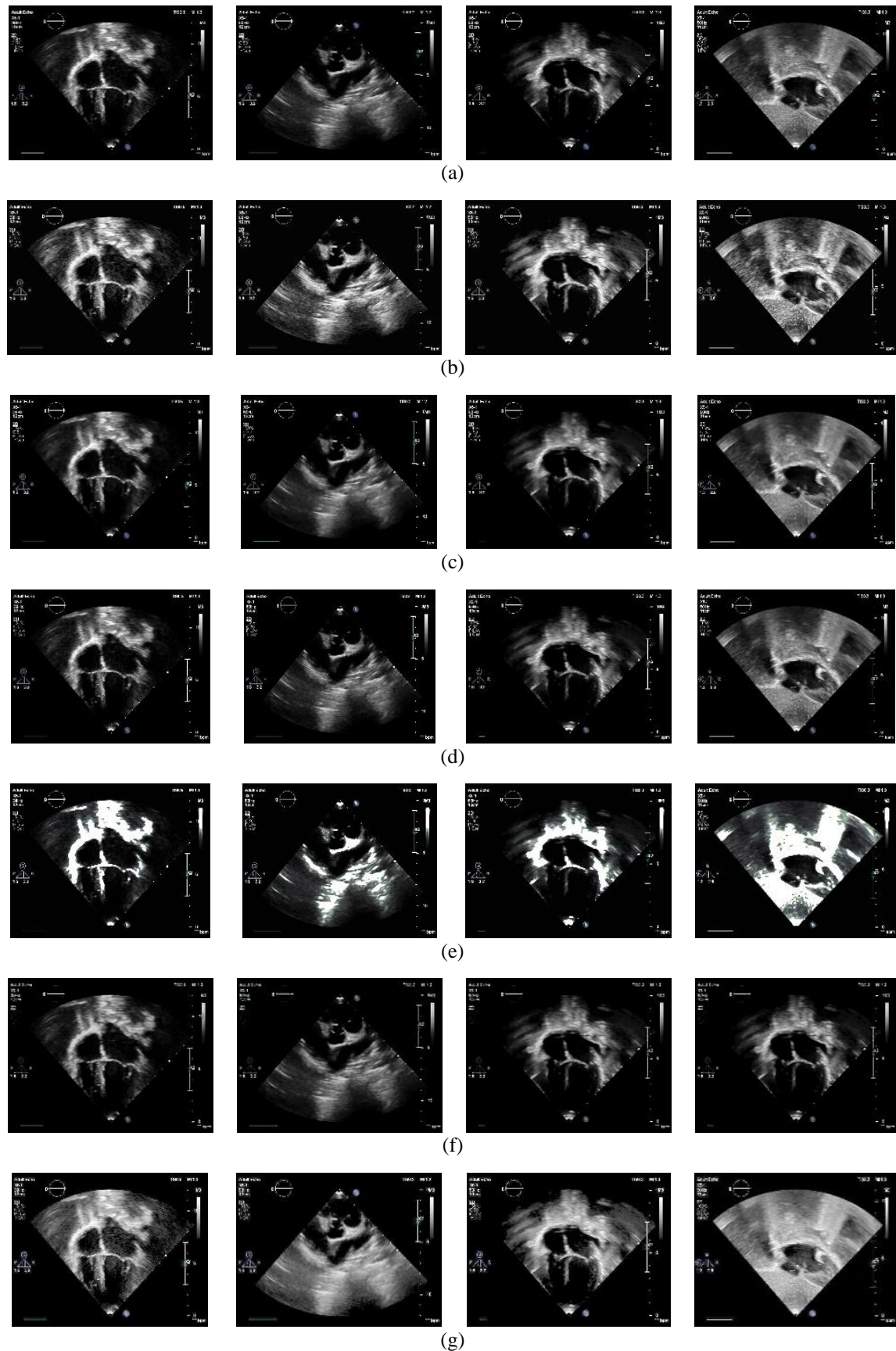


Figure 8. The sample of infant heart image enhancement with six of the existing model (a) target, (b) CLAHE, (c) bilateral filter, (d) MirNet, (e) LLCNN, (f) autoencoder, and (g) EDSR-CNNs

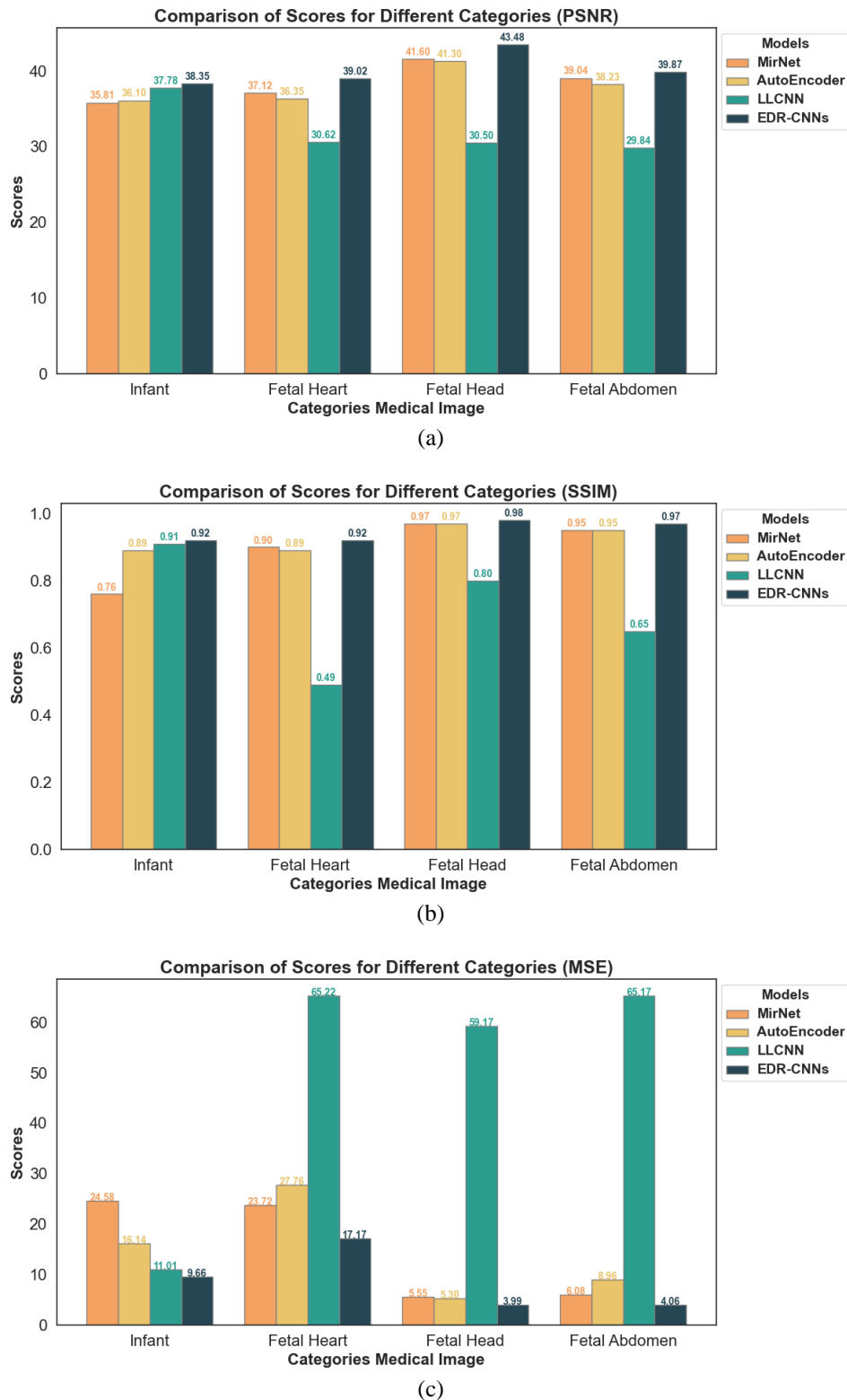


Figure 9. Performance of proposed model evaluation against other models with various datasets: (a) PSNR, (b) SSIM, and (c) MSE

We attempted to perform detection and segmentation of these fetal heart substructures using training, validation, and unseen data. The data sample is about 243 images for training data, 58 images for testing data, and 28 images for unseen data. The segmentation results before and after the image enhancement process, along with predictions using unseen data, were compared against several previously

employed models, as presented in Table 5. As indicated by the results, using the YOLACT and EDR-CNNs model produces satisfactory performance in terms of mean average precision (mAP) for bounding box (Bbox), and masking (mask). From the nine fetal heart substructures, AO, LA, MV, and IVS produce a detection rate above 60%, while LV, PV, RA, and RV produce detection in the range of 44%–55%. However, TV is hard to detect only reaching around 27%. The size of the fetal heart can vary depending on the gestational age of the fetus around 60–100 millimeters. During different stages of pregnancy, the fetal heart undergoes significant growth and development. Hence, due to the fetal heart substructure being even smaller, attaining an overlap of approximately 30%-50% Bbox and Mask between the ground truth and the predicted image proves to be quite challenging. By using unseen data indicating that the nine substructures are capable of segmenting and detecting. This suggests that the proposed model is proficient in delivering satisfactory detection performance for fetal heart substructure, both when employing validation data and unseen data, with results that outperform other methods.

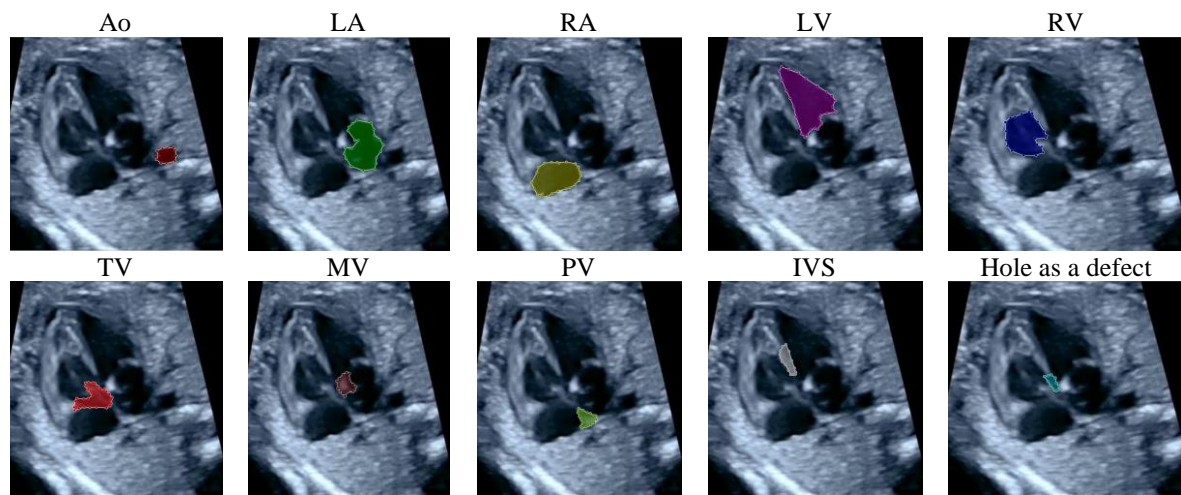


Figure 10. The sample annotation of 10 fetal heart substructures; the fetal heart has a defect in the ventricle

Table 5. The fetal heart substructure detection using EDR-CNNs image enhancement and YOLACT model with unseen data

Model		mAP (%)								
		AO	LA	RA	LV	RV	MV	IVS	TV	PV
YOLACT	BBox	46.71	52.29	37.67	32.87	43.96	26.80	55.93	15.73	42.60
	Mask	62.20	65.58	37.67	32.87	44.24	34.31	65.81	19.85	42.30
YOLACT + CLAHE	BBox	57.13	51.22	31.52	41.00	46.63	52.95	45.53	23.31	55.31
	Mask	59.13	52.37	31.19	40.08	45.99	47.78	64.14	22.44	41.40
YOLACT + Bilateral Filter	BBox	57.64	62.24	39.37	39.68	45.73	58.72	57.14	33.00	63.00
	Mask	59.52	64.24	39.21	44.12	45.42	58.32	61.23	33.22	53.14
YOLACT + MirNet	BBox	31.84	47.18	28.93	32.66	40.97	45.01	50.16	19.46	48.17
	Mask	34.75	48.39	29.98	40.43	40.01	35.07	51.22	17.54	35.21
YOLACT + LLCNN	BBox	24.32	30.98	28.34	26.22	36.84	26.30	53.29	15.92	41.11
	Mask	25.00	29.64	28.76	33.73	36.03	16.81	54.66	15.57	28.75
YOLACT + Autoencoder	BBox	34.77	45.04	32.83	32.57	40.27	50.21	58.80	22.60	51.89
	Mask	3745	45.19	32.97	38.37	39.63	46.09	58.34	17.88	46.07
YOLACT + EDR-CNN	BBox	59.08	66.60	44.97	56.53	49.66	61.96	65.90	27.25	58.58
	Mask	62.06	67.69	44.22	55.89	47.94	62.94	67.82	28.64	52.29

In this research, the proposed model was also examined for its ability to detect abnormalities in the structure of the fetal heart. In our previous investigation, our accomplishments remained constrained, yielding validation data exclusively [6], [7]. This limitation arose from the difficulty in identifying fetal hearts with unseen data, which occasionally led to instances of failure. Nevertheless, through the implementation of an enhancement process, the detection outcomes witnessed a notable advancement, reaching 60%-70% for Bbox and mask mAP, see Table 6. This marked improvement signifies a significant result and surpasses other methods.

Although the utilization of the proposed method offers numerous benefits, there are still several limitations within this research. The dataset used for comparison remains relatively restricted, encompassing only four types of medical objects, and the number of benchmarking methods is limited to five. Consideration could be given to integrating currently advanced image enhancement methods as supplementary benchmarks. Additionally, there is a dearth of definitive outcomes, such as predictions or determinations concerning the anomalies present within these medical objects. It is worth noting that research in the realm of medical imaging fundamentally serves the purpose of facilitating diagnosis. The use of the enhancement methods in terms of classification or detection of the diseases will be subject to further exploration in forthcoming investigations.

Table 6. The fetal heart abnormality as a defect detection with unseen data

Model	mAP (%)	
	BBox	Mask
YOLACT	52.79	28.65
YOLACT + CLAHE	35.43	10.33
YOLACT + Bilateral Filter	45.86	18.04
YOLACT + MirNet	58.94	42.97
YOLACT + LLCNN	51.49	43.85
YOLACT + Autoencoder	50.47	37.65
YOLACT + EDR-CNN	72.82	62.40

4. CONCLUSION

In conclusion, improving the quality of medical images, especially ultrasound images, is very important for accurate medical diagnosis and procedures. Ultrasound images often suffer from low quality due to acoustic variability, signal deficiency, attenuation, and limited resolution. These problems can interpret inhibition and medical diagnosis. To overcome this challenge, researchers have explored various methods, including image processing techniques and deep learning algorithms such as CNNs. These techniques have shown promise in reducing noise, increasing resolution, and reducing artifacts in ultrasound images, thereby increasing diagnostic accuracy. Recent advances have focused on developing specialized deep learning techniques, such as transfer learning on the proposed architecture of enhanced residual CNNs (EDR-CNNs). This architecture uses techniques such as residual learning, extended convolution, and attention-channel mechanisms to improve image quality. The results of improving the quality of the image will be tested into a special detection model to detect the structure of the fetal heart object. The results of this study can contribute to the development of technology to improve the quality of medical images in the future. The results of the detection of the structure of the fetal heart object increased from the results before the increase was carried out and after the increase had an increase that was able to increase the accuracy of the detection diagnosis. By gaining a clearer understanding of the strengths and weaknesses of various deep learning architectures, medical professionals can make informed decisions about the most effective approaches to improve the quality of medical images and consequently, improve medical diagnoses and interventions.

ACKNOWLEDGMENTS

This work was supported by Universitas Sumatera Selatan and the Intelligent Systems Research Group (ISysRG) at Universitas Sriwijaya

FUNDING INFORMATION

This research 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 Professional Research 2023 section should describe sources of funding agency that have supported the work. Authors should state how the research described in their article was funded, including grant numbers if applicable. Include the following (or similar) statement if there is no funding involved: Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ade Iriani Sapitri		✓	✓			✓			✓					
Siti Nurmaini	✓			✓					✓					✓
Muhammad Naufal		✓	✓							✓				
Rachmatullah														
Annisa Darmawahyuni						✓	✓			✓				
Firdaus Firdaus		✓							✓	✓				
Anggun Islami							✓			✓				
Bambang Tutuko		✓							✓	✓				
Akhlar Wista Arum							✓			✓				

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

All methods were conducted according to relevant guidelines and regulations. All experimental protocols were approved by Dr. Mohammad Hoesin Hospital, Indonesia. Informed consent was obtained from all subjects and/or their legal guardians.

DATA AVAILABILITY

- The code is available in <https://github.com/BestJuly/LLCNN>, while the dataset is available in <https://github.com/ISySRGg/LLCNNs>.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.




REFERENCES

- [1] J. Yu, Y. Wang, and P. Chen, "Fetal ultrasound image segmentation system and its use in fetal weight estimation," *Medical and Biological Engineering and Computing*, vol. 46, no. 12, pp. 1227–1237, Oct. 2008, doi: 10.1007/s11517-008-0407-y.
- [2] P. Vergani *et al.*, "Screening for congenital heart disease with the four-chamber view of the fetal heart," *American Journal of Obstetrics and Gynecology*, vol. 167, no. 4, pp. 1000–1003, Oct. 1992, doi: 10.1016/S0002-9378(12)80027-5.
- [3] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.
- [4] B. Al-Bander, T. Alzahrani, S. Alzahrani, B. M. Williams, and Y. Zheng, "Improving fetal head contour detection by object localisation with deep learning," in *Communications in Computer and Information Science*, vol. 1065 CCIS, Springer International Publishing, 2020, pp. 142–150, doi: 10.1007/978-3-030-39343-4_12.
- [5] S. Nurmaini *et al.*, "Deep learning-based computer-aided fetal echocardiography: Application to heart standard view segmentation for congenital heart defects detection," *Sensors*, vol. 21, no. 23, p. 8007, Nov. 2021, doi: 10.3390/s21238007.
- [6] 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.
- [7] A. Iriani Sapitri *et al.*, "Deep learning-based real time detection for cardiac objects with fetal ultrasound video," *Informatics in Medicine Unlocked*, vol. 36, p. 101150, 2023, doi: 10.1016/j.imu.2022.101150.
- [8] G. Mcleod *et al.*, "Echocardiography in congenital heart disease," *Progress in Cardiovascular Diseases*, vol. 61, no. 5–6, pp. 468–475, Nov. 2018, doi: 10.1016/j.pcad.2018.11.004.
- [9] S. Liu *et al.*, "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261–275, Apr. 2019, doi: 10.1016/j.eng.2018.11.020.
- [10] A. Paul, P. Bhattacharya, and S. P. Maity, "Histogram modification in adaptive bi-histogram equalization for contrast enhancement on digital images," *Optik*, vol. 259, p. 168899, Jun. 2022, doi: 10.1016/j.ijleo.2022.168899.
- [11] Z. Huang, Z. Wang, J. Zhang, Q. Li, and Y. Shi, "Image enhancement with the preservation of brightness and structures by employing contrast limited dynamic quadri-histogram equalization," *Optik*, vol. 226, p. 165877, Jan. 2021, doi: 10.1016/j.ijleo.2020.165877.




- [12] P. Wang, Z. Wang, D. Lv, C. Zhang, and Y. Wang, "Low illumination color image enhancement based on Gabor filtering and Retinex theory," *Multimedia Tools and Applications*, vol. 80, no. 12, pp. 17705–17719, Feb. 2021, doi: 10.1007/s11042-021-10607-7.
- [13] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, "Image enhancement for tuberculosis detection using deep learning," *IEEE Access*, vol. 8, pp. 217897–217907, 2020, doi: 10.1109/ACCESS.2020.3041867.
- [14] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Jul. 2017, vol. 2017-July, pp. 1132–1140, doi: 10.1109/CVPRW.2017.151.
- [15] X. Wang, K. Yu, S. Wu, J. Gu, and Y. Liu, "ESRGAN: Enhanced super-resolution generative adversarial networks," *arXiv:1809.00219*, pp. 1–16, 2018, doi: 10.48550/arXiv.1809.00219.
- [16] B. Kurt, V. V. Nabiye, and K. Turhan, "Medical images enhancement by using anisotropic filter and CLAHE," in *2012 International Symposium on Innovations in Intelligent Systems and Applications*, 2012, pp. 1–4, doi: 10.1109/INISTA.2012.6246971.
- [17] N. Salem, H. Malik, and A. Shams, "Medical image enhancement based on histogram algorithms," *Procedia Computer Science*, vol. 163, pp. 300–311, 2019, doi: 10.1016/j.procs.2019.12.112.
- [18] D. Bhonsle, V. Chandra, and G. R. Sinha, "Medical image denoising using bilateral filter," *International Journal of Image, Graphics and Signal Processing*, vol. 4, no. 6, pp. 36–43, Jul. 2012, doi: 10.5815/ijigsp.2012.06.06.
- [19] R. Rahman *et al.*, "Demystifying evidential Dempster Shafer-based CNN architecture for fetal plane detection from 2D ultrasound images leveraging fuzzy-contrast enhancement and explainable AI," *Ultrasonics*, vol. 132, p. 107017, Jul. 2023, doi: 10.1016/j.ultras.2023.107017.
- [20] Z. Lin *et al.*, "Quality assessment of fetal head ultrasound images based on faster R-CNN," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11042 LNCS, Springer International Publishing, 2018, pp. 38–46, doi: 10.1007/978-3-030-01045-4_5.
- [21] V. NagaPrudhviRaj and T. Venkateswarlu, "Ultrasound medical image denoising using hybrid bilateral filtering," *International Journal of Computer Applications*, vol. 56, no. 14, pp. 44–51, Oct. 2012, doi: 10.5120/8963-3171.
- [22] M. Mouzai, A. Mustapha, Z. Bousmina, I. Keskas, and F. Farhi, "Xray-Net: Self-supervised pixel stretching approach to improve low-contrast medical imaging," *Computers and Electrical Engineering*, vol. 110, p. 108859, Sep. 2023, doi: 10.1016/j.compeleceng.2023.108859.
- [23] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *IEEE International Conference on Data Mining Workshops, ICDMW*, Dec. 2016, vol. 0, pp. 241–246, doi: 10.1109/ICDMW.2016.0041.
- [24] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, and X. Xie, "LLCNN: A convolutional neural network for low-light image enhancement," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, 2017, pp. 1–4, doi: 10.1109/VCIP.2017.8305143.
- [25] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. Al Ahmad, "A review of spatial enhancement of hyperspectral remote sensing imaging techniques," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 2275–2300, 2023, doi: 10.1109/JSTARS.2023.3242048.

BIOGRAPHIES OF AUTHORS






Ade Iriani Sapitri    is currently a lecturer and a researcher with the Intelligent System Research Group, Faculty of Computer Science, Sriwijaya University, Indonesia. Her research interests include medical imaging, deep learning, and machine learning. She can be contacted at email: adeirianisapitri13@gmail.com.






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






Muhammad Naufal Rachmatullah    is currently a lecturer and a researcher at 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: naufalrachmatullah@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 by email: riset.annisadarmawahyuni@gmail.com.






Firdaus 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 systems, cardiovascular system, classification, computer aided instruction, control engineering computing, data mining, diseases, and natural 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 systems, cardiovascular systems, classification, and data mining. She can be contacted at: anggunislami2@gmail.com.



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



Akhlar Wista Arum    is currently a lecturer and researcher in Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interest includes medical imaging, biomedical signal and engineering, deep learning, and machine learning. She can be contacted at email: wistaarum16@gmail.com.