Vol. 8, No. 3, June 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

ROI-Based Shape-Prior Reconstruction for YOLOv8n-seg-Based Fetal Cerebellum Ultrasound Segmentation

Yadi Utama[1], Erwin[2], Samsuryadi[3]

[1,2,3]Computer Science Department, Sriwijaya University, Palembang, Indonesia

Received:
October 16, 2025
Revised:
May 17, 2026
Accepted:
May 28, 2026
Published:
June 22, 2026

Corresponding Author:
Author Name*:
Erwin
Email*:
erwin@unsri.ac.id

DOI:
10.63158/journalisi.v8i3.1669

Abstract. Fetal cerebellum segmentation in ultrasound images is important for quantitative analysis of fetal brain development, yet it remains challenging due to speckle noise, low contrast, acoustic artifacts, and unstable anatomical boundaries. This study proposes an ROI-Based Shape-Prior Reconstruction method as a post-processing refinement stage for YOLOv8n-seg fetal cerebellum segmentation. A total of 294 fetal ultrasound images with manually annotated binary cerebellum masks were used and divided into training, validation, and testing subsets using a 70:20:10 ratio. YOLOv8n-seg generated the initial segmentation masks, while the proposed ROI-based reconstruction stage refined the foreground region using a convolutional autoencoder trained on ROI-based binary cerebellum masks. Compared with raw YOLOv8n-seg, the proposed method improved DSC from 0.9282 to 0.9302 and IoU from 0.8671 to 0.8708. Boundary performance also improved, with HD95 decreasing from 15.06 to 14.18 and ASSD decreasing from 5.38 to 5.20. Although these improvements were modest and not statistically significant, the proposed method produced smoother boundaries and more morphologically consistent segmentation outputs in the visual evaluation. These results indicate that ROI-based shape-prior reconstruction can serve as a lightweight refinement stage for improving boundary consistency in fetal cerebellum ultrasound segmentation. However, external validation with larger datasets is still required to assess generalization.

Keywords: Fetal Ultrasound, Medical Image Segmentation, YOLOv8n-seg, Shape-Prior Reconstruction, Boundary Refinement

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Fetal ultrasound imaging is used in prenatal evaluation and monitoring of fetal brain development [1], [2]. The cerebellum is one of the important anatomical structures used to estimate gestational age and evaluate fetal neurological growth [2], [3]. Therefore, accurate cerebellum segmentation in ultrasound images is needed to support prenatal diagnosis and more reliable biometric analysis [1], [3]. In this study, the term "support" refers to assisting quantitative image analysis, not to direct clinical decision-making. However, manual cerebellum annotation still faces various challenges such as speckle noise, low contrast, acoustic shadow artifacts, and variation in results due to differences among operators in ultrasound images [1], [3]. Figure 1 shows an example of fetal cerebellum annotation in an ultrasound image.

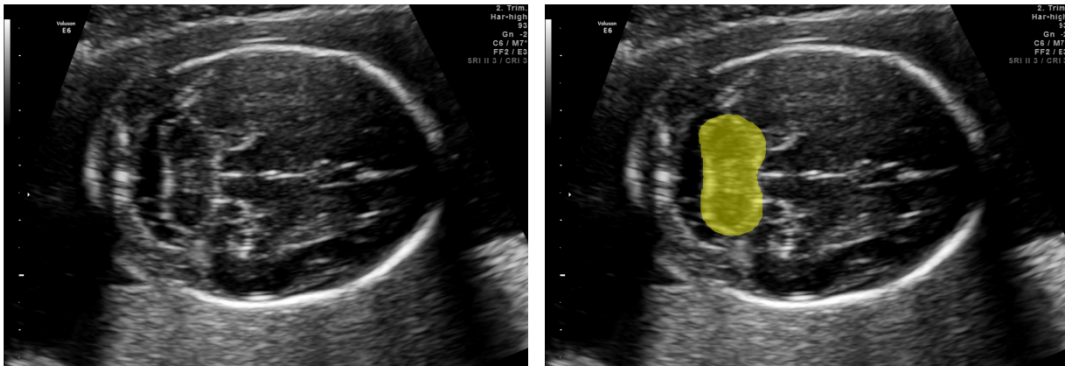

Figure 1. Example of fetal cerebellum annotation in ultrasound images

Recent developments have substantially improved medical image segmentation quality [4], [5]. CNN-based segmentation frameworks, particularly U-Net, have been widely adopted in medical image segmentation tasks [4], [6]. Several previous studies produced methods with good performance using semantic segmentation for anatomical segmentation in ultrasound images [4], [5], [7]. These methods generally produce high overlap metric values which show the effectiveness of deep learning in automatic anatomical segmentation [5], [7].

In addition to conventional encoder and decoder architectures, the YOLO model family has begun to become standard frameworks for medical image segmentation because it has high inference speed with relatively low computational complexity [8], [9]. YOLO-based segmentation models can perform object localization and pixel-level segmentation

simultaneously [8], [10]. Several recent studies have applied YOLO for organ segmentation, medical lesion segmentation, and ultrasound image analysis [9], [10]. This capability makes YOLO suitable for medical segmentation applications that require computational efficiency and fast inference [8], [9].

Several state-of-the-art ultrasound and medical image segmentation studies have reported strong performance using U-Net-based, attention-based, transformer-based, boundary-aware, shape-aware, and YOLO-based methods. These methods have been evaluated on various medical ultrasound and anatomical segmentation datasets using overlap-based metrics such as DSC and IoU, as well as boundary-based metrics such as HD95 and ASSD [5], [7], [11], [12], [13], [14], [15]. However, U-Net-based and attention-based methods can still be sensitive to speckle noise and weak boundaries, transformer-based methods generally require larger datasets and higher computational resources, and boundary-aware or shape-aware methods often require additional supervision or architectural modification. For fetal cerebellum ultrasound segmentation, prior studies have mainly improved segmentation backbones or learning strategies, while lightweight ROI-based post-processing using anatomical shape priors remains less explored. The main remaining problems include boundary instability, fragmented masks, and anatomically inconsistent shapes, especially when the cerebellum boundary is unclear due to speckle noise, low contrast, or acoustic artifacts [16], [17], [18].

High DSC and IoU values do not always indicate good anatomical consistency because small boundary distortions may not greatly change the overlapping area. Therefore, boundary-based metrics such as HD95 and ASSD are also needed to evaluate contour stability and boundary error. Although YOLOv8n-seg can efficiently generate initial fetal cerebellum masks, its output may still contain irregular boundaries or local fragmentation. Thus, an additional refinement stage is needed to improve mask morphology without modifying the YOLOv8n-seg backbone.

Previous fetal cerebellum ultrasound segmentation studies have mainly focused on supervised segmentation networks, attention-based learning, semi-supervised frameworks, or pretrained representation learning. These studies improved segmentation accuracy, but most of them still optimized the segmentation network directly and did not explicitly reconstruct the predicted cerebellum mask using an ROI-based anatomical

shape prior. Therefore, ROI-based shape-prior reconstruction for refining raw YOLOv8n-seg fetal cerebellum masks remains limited [16], [19], [20]. This gap is important because refinement can improve segmentation morphology without introducing a new architecture or retraining the YOLOv8n-seg backbone.

ROI-based shape-prior learning is suitable for fetal cerebellum segmentation because the cerebellum is a localized, compact structure in the transcerebellar plane. By focusing on the foreground ROI, the autoencoder can learn cerebellum morphology more directly while reducing background influence. This refinement aims to reduce boundary irregularities and mask fragmentation while retaining YOLOv8n-seg as the initial segmentation model. However, it is limited to improving segmentation morphology and boundary consistency, rather than establishing clinical validity or biometric measurement accuracy.

Based on these problems, this study proposes an ROI-Based Shape-Prior Reconstruction method to refine YOLOv8n-seg fetal cerebellum segmentation in ultrasound images. The method uses a convolutional autoencoder trained on ROI-based binary cerebellum masks. During inference, raw YOLOv8n-seg masks are cropped as ROIs and reconstructed using the learned shape prior to produce more consistent masks. The novelty lies in using ROI-Based Shape-Prior Reconstruction as a post-processing refinement framework, not a new YOLO architecture. This study aims to improve fetal cerebellum segmentation quality in terms of overlap accuracy and boundary consistency.

The primary contributions of this study can be summarized as follows:

1) Proposing ROI-Based Shape-Prior Reconstruction as a post-processing refinement stage for YOLOv8n-seg fetal cerebellum ultrasound segmentation.
2) Improving segmentation morphology without modifying the YOLOv8n-seg architecture.
3) Evaluating the proposed method using DSC, IoU, HD95, and ASSD.
4) Demonstrating modest, non-significant segmentation improvement over raw YOLOv8n-seg, with DSC increasing from 0.9282 to 0.9302, IoU from 0.8671 to 0.8708, HD95 decreasing from 15.06 to 14.18, and ASSD decreasing from 5.38 to 5.20.

2. RELATED WORK

Numerous studies have focused on enhancing the quality of medical image segmentation techniques. These studies generally focus on the development of segmentation architectures, improvement of feature representation, boundary refinement, and anatomical shape modeling. This section discusses related studies on deep learning-based ultrasound segmentation, YOLO-based segmentation, shape-aware segmentation, and shape-prior learning.

2.1. Deep Learning-Based Medical Image Segmentation

The development of deep learning continues to enhance the accuracy, robustness, and reliability of medical image segmentation methods [4], [5]. CNN-based architectures have increasingly become widely used approaches because they can learn spatial feature representations in complex medical images [4], [20]. U-Net is recognized as one of the most extensively utilized architectures, incorporating encoder and decoder with skip connections to maintain detailed spatial representations throughout the feature reconstruction stage [5], [20].

Various developments of U-Net have been carried out to achieve more accurate and reliable medical image segmentation [4], [5]. Attention U-Net improves segmentation performance by directing the feature extraction process toward anatomically significant regions while minimizing the influence of irrelevant background information [5], [11]. U-Net++ introduces dense skip connections to improve multi-scale feature integration between the encoder and decoder [21]. Then, DeepLabV3+ improves feature extraction through atrous spatial pyramid pooling to enhance contextual feature extraction at various scales [22].

In addition to CNN-based architectures, transformer-based approaches have also begun to be widely adopted for medical image segmentation tasks [4], [12], [21]. Transformer allows the model to learn global relationships between image areas through the self-attention mechanism, so it can capture broader context than conventional CNN [12], [21]. Several studies reported that the combination of CNN and transformer can improve segmentation performance in various medical applications [7], [12], [23].

Although these various methods produce high overlap metrics, most of them still focus on pixel-level optimization without explicitly maintaining anatomical shape consistency [13], [24]. Several issues related to ultrasound image quality can cause unstable boundaries and inconsistent anatomical shapes [11], [18]. Thus, high DSC and IoU values do not always indicate good anatomical morphology quality in the segmentation results [11], [13], [24].

2.2. YOLO-Based Medical Image Segmentation

In addition to conventional encoder-decoder architectures, YOLO-based models, You Only Look Once, have also begun to be used in medical image segmentation because they have high inference speed with relatively low computational complexity [8], [9]. Previously, YOLO was developed for real-time object detection, but recent developments allow the integration of pixel-level segmentation in one unified framework [8], [10]. This approach allows object localization and segmentation to be performed simultaneously, so it is more efficient than conventional segmentation approaches that require separate stages [8], [9], [10].

Various YOLO variants have been applied to the segmentation of organs, lesions, tumors, and anatomical regions in medical imaging data, including ultrasound images [8], [9], [10]. YOLO-based segmentation has advantages in computational efficiency, inference speed, and generalization ability for various object sizes [8], [25]. The relatively lightweight YOLO architecture is also suitable for real-time implementation and devices with limited computational resources [8], [25].

YOLOv8 is one of the latest generations of the YOLO family that supports segmentation natively through the integration of a detection head and a segmentation head in one model [8], [25]. YOLOv8 shows competitive segmentation performance with high inference efficiency [8], [9]. In medical ultrasound images, the ability to localize the region of interest (ROI) quickly makes YOLOv8 suitable for segmenting small anatomical structures with complex boundaries [9], [10], [26].

Nevertheless, YOLO-based segmentation still has several limitations in ultrasound images [8], [9]. The segmentation results are also still often influenced by noise, boundary fragmentation, discontinuous areas, and anatomical shape deformation due to low image quality [17], [18], [20]. Most YOLO-based segmentation studies still focus on feature

extraction, attention mechanism, and loss function optimization, while anatomical shape-prior learning for segmentation refinement is still limited [8], [13], [26]. Therefore, an additional approach is needed to improve anatomical morphology consistency in YOLO-based fetal cerebellum segmentation in ultrasound images [13], [20].

2.3. Shape-Aware and Boundary-Aware Segmentation

To mitigate the inherent limitations of pixel-based segmentation, various studies have developed shape-aware and boundary-aware segmentation approaches to improve anatomical morphology quality [11], [13]. These approaches aim to maintain boundary continuity, smooth the segmentation shape, and reduce anatomical deformation in medical images with high noise [11], [13], [17], [18].

Several studies introduced boundary enhancement modules to improve the sharpness of segmentation edges through explicit boundary feature learning [11], [14]. Other approaches use contour-aware learning, edge-guided attention, and boundary loss to enhance the model's responsiveness toward anatomically meaningful boundary representations [14], [27]. In addition, several methods combine segmentation networks with active contour models, level-set refinement, or a Conditional Random Field (CRF)-based post-processing strategy to produce smoother and more consistent segmentation boundaries [28].

On the other hand, shape-aware segmentation has begun to be developed by using shape constraints and anatomical priors during the training process [13], [15]. This approach aims to produce segmentation that remains consistent with valid anatomical structures [13], [15]. Several studies use statistical shape models, latent representation learning, and adversarial learning to learn the distribution of anatomical shapes in medical data [15], [29]. With shape constraints, the model is expected to reduce segmentation results that are morphologically unrealistic [13], [15].

Although these various approaches improve boundary quality and anatomical shape consistency, most of them still require complex supervised learning mechanisms and additional annotations or constraints during training [13], [15], [20]. In many cases, refinement also still depends on feature representation from the segmentation backbone without specifically learning anatomical shape-prior [8], [15], [30]. Therefore, a simpler but

still effective approach is needed to improve anatomical shape consistency in YOLO-based ultrasound segmentation [20], [30].

2.4. Shape-Prior Learning

Shape-prior learning is a developing approach in medical image analysis to learn anatomical shape representation [31]. This approach generally uses autoencoder or latent representation learning to learn the morphological distribution of objects based on binary segmentation masks [15], [31]. By learning shape representation in the latent space, the model can reconstruct more consistent anatomical structures and reduce shape deformation in the segmentation results [15], [31].

Autoencoder is widely used in shape-prior learning because it can compress anatomical shapes into low-dimensional latent representations and reconstruct them again through the decoding process [15], [31]. This approach is used to learn the structures of organs, blood vessels, brain tissues, and other biological structures [31], [32]. The learned shape-prior representation allows the model to understand global morphological patterns so that it demonstrates greater robustness against noise-induced variations and local variations in images [15], [31], [32].

Several studies have integrated shape-prior learning with segmentation networks to improve the anatomical consistency of segmentation results [13], [15], [31]. However, most approaches still focus on conventional encoder-decoder-based segmentation and have not been widely applied to YOLO-based segmentation [8], [15], [33]. In addition, the majority of methods still use supervised or semi-supervised learning that requires additional constraints during the training process [13], [20]. Meanwhile, fetal cerebellum segmentation in ultrasound images with high speckle noise, unclear boundaries, and shape variation due to fetal position makes anatomical shape learning more difficult than other medical image modalities [18], [20].

Based on these limitations, this study proposes an ROI-Based Shape-Prior Reconstruction approach for YOLO-based fetal cerebellum segmentation refinement. This study uses a convolutional autoencoder to learn the anatomical shape-prior of the cerebellum from ROI-based cropped binary masks. The learned shape-prior is then used to reconstruct

and improve YOLOv8n-seg segmentation results in the inference stage, producing anatomically plausible segmentation.

3. PROPOSED METHOD

This study proposes an ROI-Based Shape-Prior Reconstruction method to improve the morphological consistency of YOLOv8n-seg segmentation for fetal cerebellum ultrasound images. The overall workflow is shown in Figure 2. The proposed method consists of two stages: initial segmentation using YOLOv8n-seg and shape-prior reconstruction using a convolutional autoencoder.

In the first stage, YOLOv8n-seg generates an initial binary cerebellum mask from the input ultrasound image. In the second stage, the raw mask is used to localize the cerebellum ROI, which is cropped with padding, resized to a fixed size, and reconstructed by the autoencoder to obtain a more anatomically consistent mask. The autoencoder was trained using ROI masks generated only from the training subset, while validation ROI masks were used only for monitoring and testing ROI masks were used only for final evaluation to prevent data leakage.

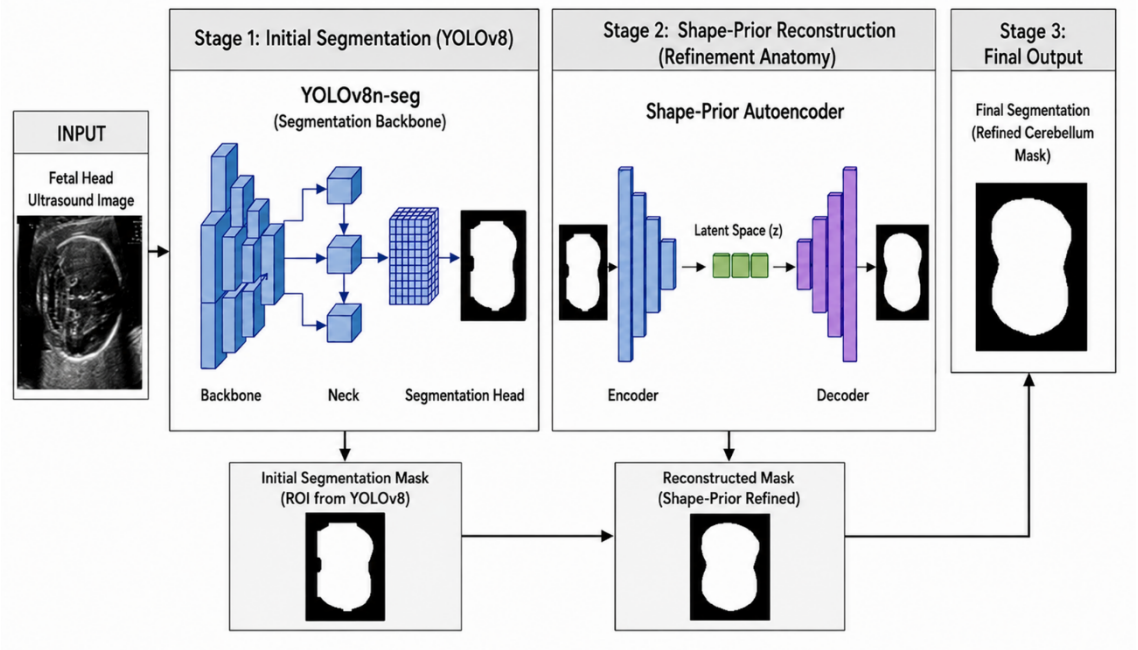

Figure 2. Workflow of the proposed ROI-Based Shape-Prior Reconstruction method

During inference, the raw YOLOv8n-seg mask is cropped as an ROI, reconstructed by the trained autoencoder, resized, and restored to its original position to produce a refined cerebellum mask with improved boundary consistency and anatomical shape stability, without modifying the YOLOv8n-seg architecture.

3.1. Dataset Preparation

The dataset was obtained from a public Zenodo fetal head ultrasound repository [34]. A total of 294 transcerebellar-plane images with visible cerebellum regions were manually selected and annotated in LabelMe as binary masks, with the cerebellum as foreground and the remaining area as background. Low-quality images, unclear boundaries, and incomplete cerebellum appearances were excluded. Annotation was performed by one trained annotator experienced in fetal ultrasound image annotation and reviewed for mask completeness, boundary consistency, foreground-background labeling accuracy, and correspondence with the visible cerebellum region. The annotations were checked once after completion and were not independently repeated by additional annotators. No independent clinical adjudication or multi-annotator consensus was performed, which is acknowledged as a study limitation.

The dataset was image-level split into 206 training, 59 validation, and 29 testing images using a 70:20:10 ratio, as shown in Table 1. Patient-level splitting was not possible because patient identifiers were unavailable, so potential patient-level overlap could not be fully excluded. The testing subset was excluded from all training processes, with random seed 42 fixed for reproducibility. Images and masks were resized to 512 × 512 pixels for YOLOv8n-seg training, with augmentation applied only to training data using horizontal flipping, ±10° rotation, translation, and scaling. For the shape-prior autoencoder, cerebellum foreground ROIs were cropped with 20-pixel padding and resized to 128 × 128 pixels using nearest-neighbor interpolation.

Table 1. Dataset distribution used in this study

Subset	Number of Images	Percentage (%)
Training	206	70
Validation	59	20
Testing	29	10

3.2. YOLO-Based Fetal Cerebellum Segmentation

YOLOv8n-seg was used as the segmentation backbone to generate the initial fetal cerebellum mask. This model was selected because it performs object localization and pixel-level segmentation simultaneously with efficient inference and relatively low computational complexity.

A pretrained YOLOv8n-seg model was fine-tuned using transfer learning to improve convergence and training stability on the limited ultrasound dataset. All input images were resized to 512 × 512 pixels. During inference, the raw binary cerebellum mask produced by YOLOv8n-seg was used as the input for the subsequent shape-prior reconstruction stage.

3.3. ROI-Based Shape-Prior Reconstruction

To improve the morphological consistency of ultrasound segmentation results, this study proposes an ROI-Based Shape-Prior Reconstruction method using a convolutional autoencoder. This method learns the fetal cerebellum shape prior through binary mask reconstruction.

The proposed autoencoder consists of an encoder, latent space, and decoder. The input is a 128 × 128 single-channel binary ROI mask. The encoder uses four convolutional blocks, each consisting of 3 × 3 convolution, batch normalization, ReLU activation, and 2 × 2 max-pooling. The channels increase from 1 to 16, 32, 64, and 96, while the spatial size is reduced from 128 × 128 to 8 × 8.

The encoded feature map is flattened into a 16-dimensional latent vector that represents the compact cerebellum shape prior. The decoder reconstructs the mask by projecting this vector back to an 8 × 8 × 96 feature map, followed by four upsampling blocks with 2 × 2 upsampling, 3 × 3 convolution, batch normalization, and ReLU activation. The channels are reduced from 96 to 64, 32, 16, and 1. A sigmoid layer produces the probability map, which is binarized using a 0.5 threshold.

Before training, ground-truth cerebellum masks from the training subset were cropped using a foreground bounding box with 20-pixel padding and resized to 128 × 128 using nearest-neighbor interpolation. The same ROI extraction was applied to validation and

testing data only for monitoring and final evaluation, and no testing ROI masks were used for training. The 16-dimensional latent space was used to obtain a compact anatomical representation, while the 20-pixel padding preserved boundary context during cropping.

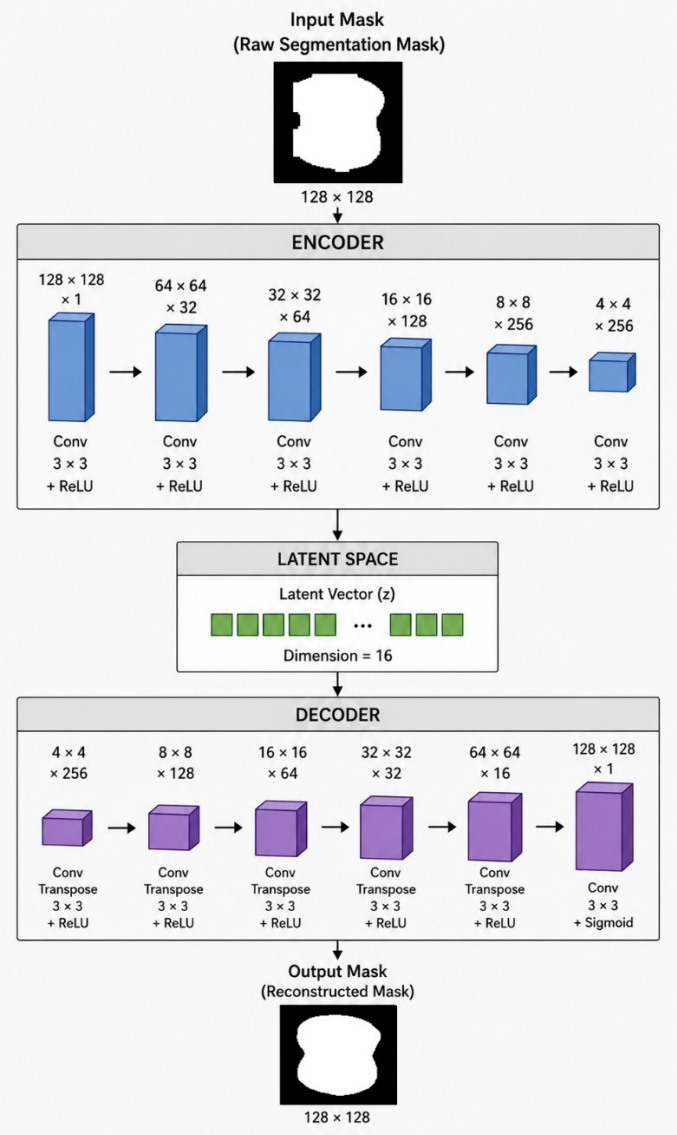

Figure 3. Architecture of the proposed convolutional autoencoder

In the encoding stage, the input binary mask x is mapped into a latent representation:

$$z = E(x) \tag{1}$$

where x is the input binary cerebellum mask, E(x) is the encoder function, and z is the latent vector.

The latent vector is then decoded to reconstruct the cerebellum mask:

$$M_{prob} = D(z) \tag{2}$$

where D(z) is the decoder function, and M_{prob} is the reconstructed probability map. Since the decoder output uses sigmoid activation, each pixel value ranges from 0 to 1.

The reconstructed probability map is then converted into a binary mask using a threshold of 0.5. Pixels with values greater than or equal to 0.5 are assigned as cerebellum foreground, while pixels below 0.5 are assigned as background. This thresholding process is written as:

$$M_{bin}(i) = 1, \text{if } M_{prob}(i) \geq 0.5, \text{and } 0 \text{ otherwise} \tag{3}$$

where $M_{prob}(i)$ is the probability value of pixel i, and $M_{bin}(i)$ is the final binary reconstructed mask.

To maintain the reconstruction quality of the anatomical shape, the proposed model was optimized using BCE-Dice loss. BCE loss is used to optimize pixel-wise reconstruction accuracy [35]:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[x_i \cdot \log\left(M_{prob}(i)\right) + (1 - x_i) \cdot \log(1 - M_{prob}(i))] \tag{4}$$

where x_i denotes the ground-truth pixel value, $M_{prob}(i)$ is the reconstructed probability value, and N is the total number of pixels.

Dice loss is used to preserve anatomical overlap consistency [35]:

$$L_{Dice} = 1 - \frac{2 \cdot \sum_{i=1}^{N} x_i \cdot M_{prob}(i) + \epsilon}{\sum_{i=1}^{N} x_i + \sum_{i=1}^{N} M_{prob}(i) + \epsilon} \tag{5}$$

where ϵ is a small constant used to avoid division by zero. The total loss is defined as:

$$L_{Total} = L_{BCE} + L_{Dice} \tag{6}$$

The combination of BCE loss and Dice loss allows the autoencoder to maintain pixel-level reconstruction accuracy and anatomical shape consistency. By learning the cerebellum shape distribution in the latent space, the autoencoder produces a reconstructed mask that is more morphologically stable and less sensitive to local boundary noise.

3.4. ROI-Based Shape-Prior Reconstruction During Inference

The inference stage integrates YOLOv8n-seg-based initial segmentation with shape-prior reconstruction to produce an anatomically stable refined segmentation mask. The reconstruction flow is shown in Figure 4.

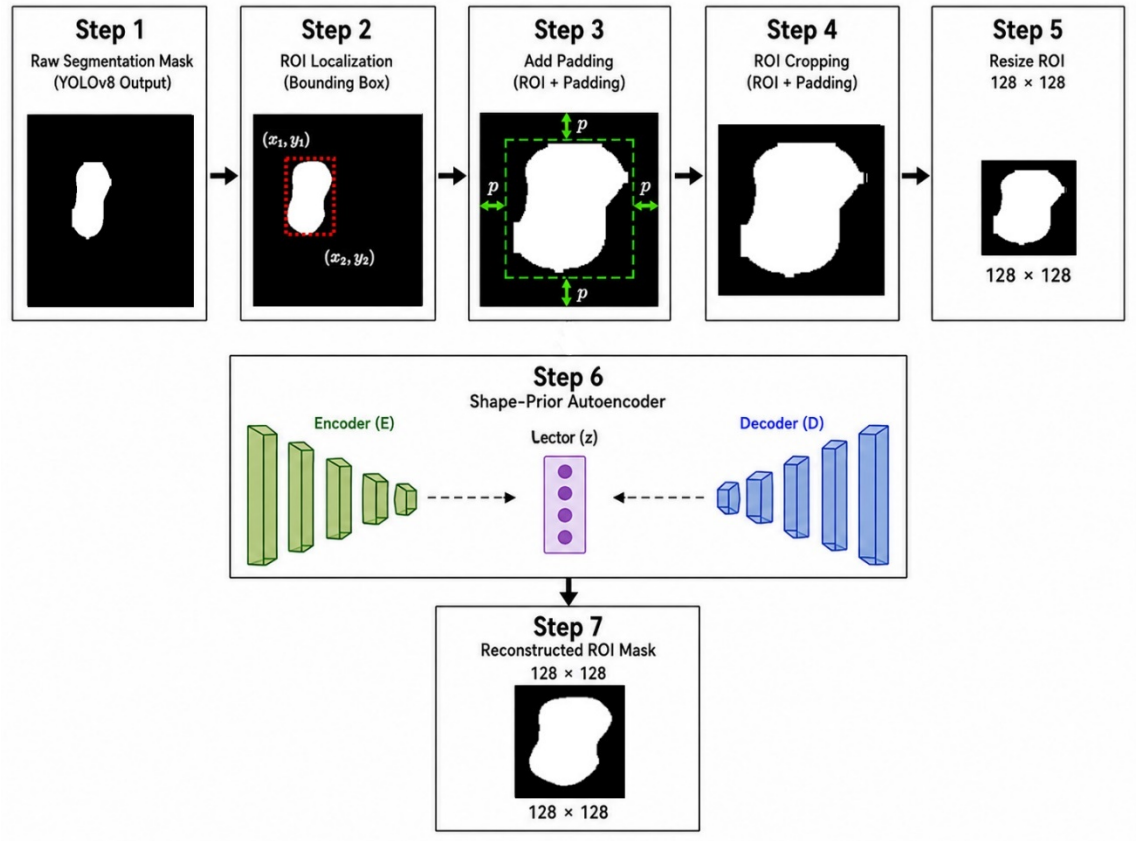

Figure 4. ROI-based shape-prior reconstruction process during inference

In Step 1, the ultrasound image is processed using YOLOv8n-seg to produce the raw cerebellum segmentation mask. The initial output is a binary mask that represents the predicted cerebellum foreground area.

In Step 2, ROI localization is performed using a foreground mask-based bounding box. The bounding box is determined from the foreground pixel coordinates of the raw segmentation mask and is expressed as:

$$BBox = (x_1, y_1, x_2, y_2) \tag{7}$$

where BBox denotes the ROI bounding box, (x_1, y_1) is the upper-left coordinate, and (x_2, y_2) is the lower-right coordinate.

If the YOLOv8n-seg output produced an empty mask or no foreground pixels were detected, ROI extraction was not performed and the image was treated as a failed segmentation case. For poor or noisy masks, the largest connected component was retained before bounding box extraction, while very small components were removed to avoid unstable ROI generation. This procedure ensured that the autoencoder received a valid ROI mask for shape-prior reconstruction.

In Step 3, the bounding box is expanded using 20-pixel padding to preserve anatomical context around the cerebellum. The padded bounding box is written as:

$$BBox_{padding} = (x_1 - p, y_1 - p, x_2 + p, y_2 + p) \tag{8}$$

where p is the number of padding pixels, and $BBox_{padding}$ is the bounding box after padding addition.

In Step 4, ROI cropping is performed based on the padded bounding box. The cropped ROI mask is then resized to 128 × 128 pixels in Step 5 using nearest-neighbor interpolation to preserve binary values.

In Step 6, the resized ROI mask is entered into the shape-prior autoencoder. The reconstruction process is written as:

$$M_{prob} = Dec(Enc(M_{ROI})) \quad (9)$$

where M_{ROI} is the ROI mask from YOLOv8n-seg segmentation, Enc is the encoder, Dec is the decoder, and M_{prob} is the reconstructed probability map.

The reconstructed probability map is converted into a binary mask using the same threshold of 0.5:

$$M_{bin}(i) = 1, \text{if } M_{prob}(i) \geq 0.5, \text{and } 0 \text{ otherwise} \quad (10)$$

where $M_{bin}(i)$ is the binary reconstructed ROI mask at pixel i.

In Step 7, the binary reconstructed ROI mask is resized back to its original ROI size and placed into its original position in the full-size image canvas. This reconstructed mask is used as the refined segmentation result. Compared with the raw YOLOv8n-seg output, the reconstructed result has a more stable boundary and better anatomical shape consistency.

The proposed approach performs reconstruction based on the anatomical shape distribution learned in the latent space. Therefore, the final segmentation result has a more stable boundary while maintaining the anatomical shape consistency of the cerebellum in ultrasound images.

3.5. Experimental Setup

All experiments were implemented in Python using PyTorch and Ultralytics YOLOv8, and were accelerated using a CUDA-enabled NVIDIA GeForce RTX 4070 GPU. The software environment included Python 3.13.2, PyTorch 2.5.1, and Ultralytics 8.3.102. A fixed random seed of 42 was used for dataset splitting and training configuration to support reproducibility.

In the segmentation stage, a pretrained YOLOv8n-seg was used as the backbone. All images were resized to 512 × 512 pixels. The model was trained for 300 epochs with a batch size of 16 using AdamW optimizer, a learning rate of 1×10^{-4}, weight decay of 1×10^{-4}, and a cosine learning rate scheduler.

In the ROI-Based Shape-Prior Reconstruction stage, the convolutional autoencoder was trained using ROI binary cerebellum masks generated from foreground mask-based cropping. To prevent data leakage, only ROI masks from the training subset were used for training, while validation data were used only for monitoring and testing data only for final evaluation. All ROI masks were resized to 128 × 128 pixels. The autoencoder was trained for 200 epochs with a batch size of 16 using Adam optimizer, a learning rate of 1×10^{-3}, and hybrid BCE–Dice loss.

A latent dimension of 16 was used to balance compact shape representation and stable reconstruction, preserving global cerebellum morphology while reducing pixel-level noise. A 20-pixel ROI padding preserved boundary context and minimized contour truncation during cropping, especially for irregular raw YOLOv8n-seg masks.

All main training hyperparameters, including image size, batch size, learning rate, optimizer, number of epochs, latent dimension, ROI size, and ROI padding, were fixed before the final evaluation and were not extensively tuned on the testing subset. Segmentation metrics were calculated using Python-based implementations with NumPy 2.0.2, OpenCV 4.13.0, and SciPy 1.16.3 distance-transform functions.

3.6. Evaluation Metrics

This study employs overlap-based and boundary-based metrics to assess the segmentation quality.

Dice Similarity Coefficient (DSC) measures the overlap between the segmentation mask and the ground truth mask [36]:

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \tag{11}$$

where P denotes the predicted segmentation area and G denotes the ground truth mask.

Intersection over Union (IoU) measures the ratio between the intersection and union areas of the prediction and ground truth [36]:

$$IoU = \frac{|P \cap G|}{|P \cup G|} \quad (12)$$

Hausdorff Distance 95% (HD95) evaluates boundary accuracy using the 95th-percentile Hausdorff distance between the predicted and ground-truth contours [36]:

$$HD_{95}(P, G) = \max\left\{percentile_{95}\big(d(P, G)\big), percentile_{95}(d(G, P))\right\} \quad (13)$$

where d(P, G) represents the boundary distance between the prediction result and the ground truth.

Average Symmetric Surface Distance (ASSD) measures the average symmetric distance between the predicted and ground-truth boundaries [37]:

$$ASSD(P, G) = \frac{\sum_{p \epsilon P} d(p,G) + \sum_{g \epsilon G} d(g,P)}{|P| + |G|} \quad (14)$$

mAP50 and mAP50-95 are used to evaluate the YOLOv8n-seg segmentation performance. mAP50 measures average precision at an IoU threshold of 0.5, while mAP50-95 computes the mean average precision across IoU thresholds from 0.5 to 0.95 [8], [9].

4. RESULTS AND DISCUSSION

4.1. Training Performance Analysis

In this study, the training process was conducted in two main stages, namely training the YOLOv8n-seg segmentation model as the initial segmentation and training the convolutional autoencoder for ROI-Based Shape-Prior Reconstruction. Training performance evaluation was performed to analyze the convergence stability of the model, segmentation quality, and the ability to learn shape-prior representations during the training process.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

4.1.1. YOLOv8n-seg Training Performance

Figure 5 presents the YOLOv8n-seg training curves, including training segmentation loss, validation mAP50-95, and the learning rate schedule. The segmentation loss decreased steadily during training, while validation mAP50-95 remained stable in the final epochs, indicating consistent segmentation performance on the validation data. The cosine learning rate scheduler also supported stable optimization and model convergence.

The best checkpoint was selected based on validation Dice score to ensure optimal anatomical segmentation quality during inference. As shown in Figure 6, validation Dice increased in the early training stage and then stabilized. The best checkpoint was obtained at epoch 64 with a Dice score of 0.9321 and was used for all inference and evaluation processes.

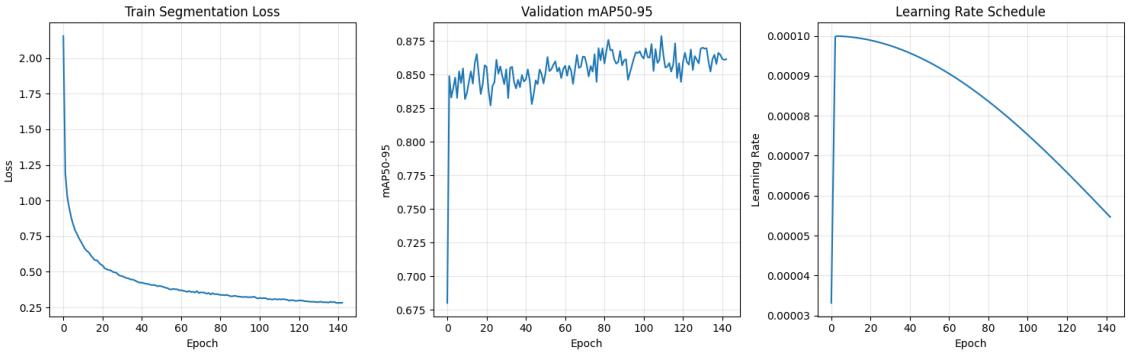

Figure 5. Training performance curves of YOLOv8n-seg

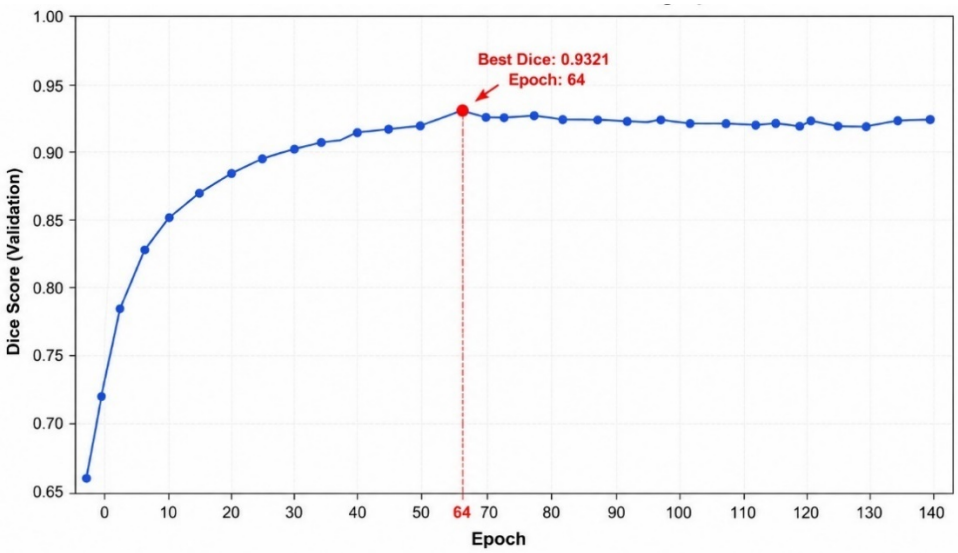

Figure 6. Validation Dice-based checkpoint selection during YOLOv8n-seg training

4.1.2. ROI-Based Shape-Prior Reconstruction Training Performance

In addition to YOLOv8n-seg training, this study also trained ROI-Based Shape-Prior Reconstruction using a convolutional autoencoder to learn the anatomical shape distribution of the cerebellum from ROI binary masks. The learning process was performed using anatomical mask representations obtained from foreground region-based cropping. Training evaluation was performed to analyze model convergence stability and anatomical shape reconstruction quality on validation data.

Figure 7 shows the training and validation loss curves of the reconstruction module. Both losses decreased and stabilized with a small gap, indicating stable convergence, consistent cerebellum shape learning, and no significant overfitting during ROI-based morphology reconstruction.

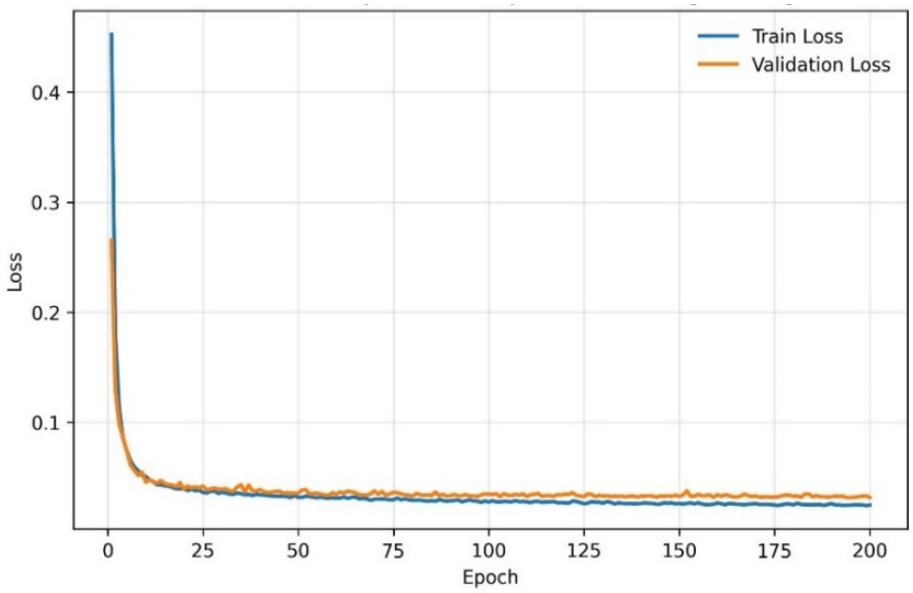

Figure 7. Training loss curves of the proposed shape-prior reconstruction autoencoder

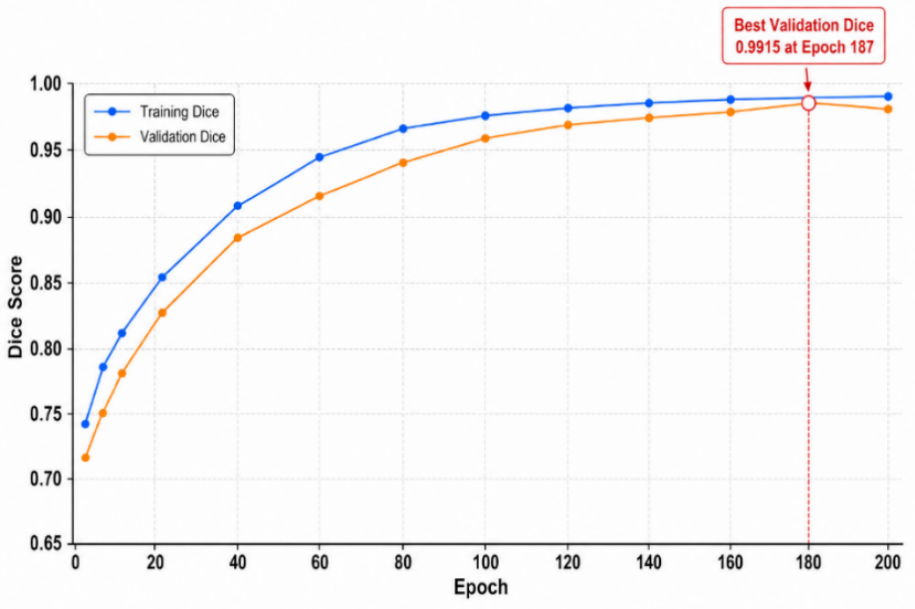

Figure 8. Training and validation Dice performance during ROI-Based Shape-Prior Reconstruction

Anatomical reconstruction performance was evaluated using Dice during training. As shown in Figure 8, training and validation Dice increased consistently and stabilized in the final stage, indicating that the autoencoder maintained cerebellum shape consistency during mask reconstruction. The best validation Dice reached 0.9915, with a validation IoU of 0.9831, showing very high morphological reconstruction quality. Table 2 summarizes the best ROI-Based Shape-Prior Reconstruction performance. The model achieved a validation Dice of 0.9915, validation IoU of 0.9831, and final validation loss of 0.0316, indicating high anatomical reconstruction accuracy, low reconstruction error, and stable optimization. The combined BCE-Dice loss also helped balance pixel-level accuracy and anatomical structure consistency during learning.

To further evaluate the reconstruction capability of the proposed autoencoder, test-set ROI mask reconstruction was also assessed independently using ground-truth ROI masks. In this evaluation, the ground-truth cerebellum mask from each test image was cropped using the foreground ROI, resized to 128 × 128 pixels, reconstructed by the trained autoencoder, and compared with the corresponding normalized ground-truth ROI mask. As shown in Table 3, the autoencoder achieved a test reconstruction Dice of 0.9869 ± 0.0050 and an IoU of 0.9743 ± 0.0096, with low boundary errors indicated by an HD95 of 1.31 ± 0.46 and an ASSD of 0.55 ± 0.15. These results indicate that the autoencoder can

reconstruct normalized fetal cerebellum ROI masks with high morphological consistency. However, this performance reflects reconstruction quality in ROI space, while the final full-image segmentation performance still depends on YOLOv8n-seg mask quality, ROI localization, and paste-back alignment.

Table 2. Best training performance of ROI-Based Shape-Prior Reconstruction

Metric	Best Value
Validation Dice	0.9915
Validation IoU	0.9831
Final Validation Loss	0.0316

Table 3. Test-set reconstruction performance of the ROI-Based Shape-Prior Reconstruction autoencoder

Metric	Test reconstruction performance (mean ± std)
DSC	0.9869 ± 0.0050
IoU	0.9743 ± 0.0096
HD95	1.31 ± 0.46
ASSD	0.55 ± 0.15

4.2. Quantitative Segmentation Performance

Segmentation performance was evaluated by comparing raw YOLOv8n-seg and reconstructed masks using overlap- and boundary-based metrics. Per-image mean ± standard deviation was reported, followed by paired statistical analysis using the Shapiro-Wilk normality test and either paired t-test or Wilcoxon signed-rank test. In addition, per-image distributions of DSC, IoU, HD95, and ASSD on the testing subset were analyzed using boxplots to illustrate performance variability across test images, as shown in Figure 9. Inference efficiency was also evaluated using inference time in milliseconds per image and frames per second (FPS) to assess the computational impact of the reconstruction stage.

Based on Table 4, the proposed method slightly improved YOLOv8n-seg performance, with higher DSC and IoU and lower HD95 and ASSD, indicating better overlap and

boundary consistency without modifying the YOLOv8n-seg architecture. In addition, based on YOLO validation, mAP50 remained stable, while mAP50-95 increased from 0.8538 to 0.8600. However, paired statistical analysis showed that these improvements were not statistically significant for DSC, IoU, HD95, and ASSD, while inference time increased from 16.08 ± 8.43 ms/image to 24.16 ± 13.18 ms/image, corresponding to an additional processing cost of 8.09 ms/image. Consistently, FPS decreased from 68.25 ± 12.93 to 46.03 ± 10.23 due to the additional reconstruction stage. To quantify this cost, the ROI-Based Shape-Prior Reconstruction module was evaluated using parameter count, MACs, and FLOPs; since the autoencoder processes only a compact 1 × 128 × 128 ROI binary mask rather than the full ultrasound image, the additional computational burden remains limited to localized reconstruction.

Table 4. Per-image quantitative comparison and paired statistical analysis on the testing subset

Metric	YOLOv8n-seg (mean ± std)	Proposed Method (mean ± std)	Mean Difference	95% CI	p-value
DSC	0.9282 ± 0.0256	0.9302 ± 0.0285	+0.0020	[-0.0012, 0.0052]	0.2160
IoU	0.8671 ± 0.0436	0.8708 ± 0.0487	+0.0037	[-0.0017, 0.0092]	0.1734
HD95	15.06 ± 6.84	14.18 ± 6.11	−0.88	[-2.01, 0.25]	0.1221
ASSD	5.38 ± 2.30	5.20 ± 2.32	−0.18	[-0.45, 0.09]	0.1835
Inference Time (ms/image)	16.08 ± 8.43	24.16 ± 13.18	+8.09	[5.84, 10.33]	<0.0001
FPS	68.25 ± 12.93	46.03 ± 10.23	−22.22	[-26.04, -18.40]	<0.0001

Table 5. Computational complexity of the ROI-Based Shape-Prior Reconstruction module

Component	Input Size	Total Parameters	Trainable Parameters	MACs	FLOPs
Autoencoder	1 × 128 × 128	770,113	770,113	422.23 M	844.46 M

Table 5 shows that the ROI-Based Shape-Prior Reconstruction autoencoder has 770,113 trainable parameters, or about 0.77M, with 422.23M MACs and 844.46M FLOPs for a 1 × 128 × 128 ROI binary mask input. This indicates that the module is lightweight, with localized and manageable computational cost because refinement is performed only at the ROI level rather than on the full ultrasound image. Although the improvements were

not statistically significant, they remain practically relevant for boundary refinement, as the method mainly corrects local contour irregularities. These changes are more evident in HD95 and ASSD than in DSC and IoU, because overlap metrics are dominated by foreground area, while boundary metrics are more sensitive to contour displacement.

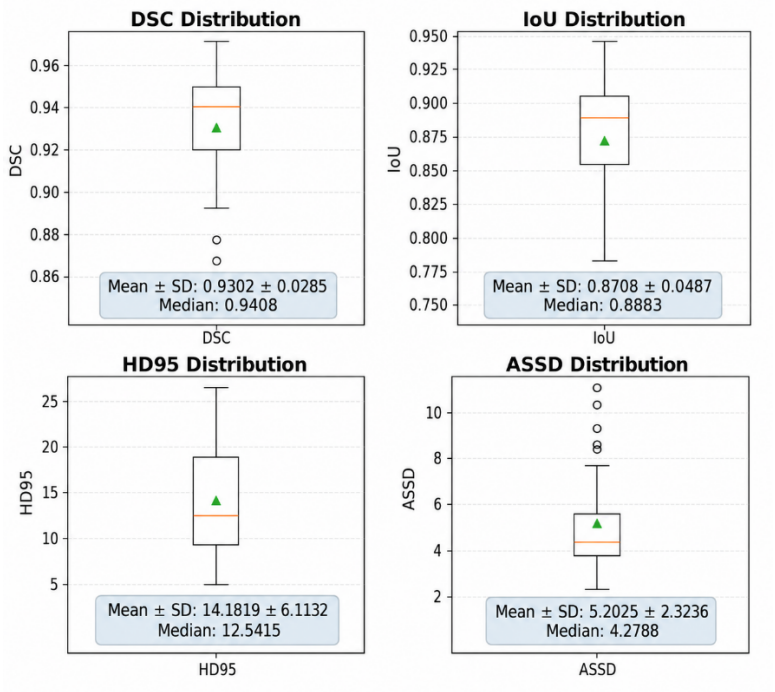

Figure 9. Boxplot distribution of per-image DSC, IoU, HD95, and ASSD on the testing subset

The boxplots show the per-image distribution of overlap- and boundary-based metrics across the test images. DSC and IoU represent segmentation overlap variability, while HD95 and ASSD reflect boundary error variability. Overall, the reconstructed masks show a distribution pattern similar to raw YOLOv8n-seg outputs, with slight improvements in overlap and boundary metrics across several cases.

4.3. Ablation Study

An ablation study was conducted to evaluate the contribution of each component. YOLOv8n-seg was used as the baseline with three variants: ROI cropping only, autoencoder without ROI, and the full proposed method combining ROI localization and shape-prior reconstruction.

Table 6. Ablation study of the proposed reconstruction components

Method	DSC	IoU	HD95	ASSD	Inference Time (ms/image)
YOLOv8n-seg	0.9282 ± 0.0256	0.8671 ± 0.0436	15.06 ± 6.84	5.38 ± 2.30	16.08 ± 8.43
YOLOv8n-seg + ROI cropping (resizing only)	0.9286 ± 0.0281	0.8679 ± 0.0478	14.75 ± 6.23	5.23 ± 2.21	19.78 ± 9.80
YOLOv8n-seg + Autoencoder	0.9151 ± 0.0294	0.8448 ± 0.0487	48.52 ± 128.13	10.46 ± 18.34	21.74 ± 8.77
Full Proposed Method	**0.9302 ± 0.0285**	**0.8708 ± 0.0487**	**14.18 ± 6.11**	**5.20 ± 2.32**	**24.16 ± 13.18**

The ROI cropping-only variant was used to verify whether improvement came merely from cropping and resizing rather than reconstruction. As shown in Table 6, the full proposed method achieved the best overall performance, especially in boundary-based metrics, indicating that ROI localization and autoencoder-based shape-prior reconstruction jointly improve morphological consistency.

The autoencoder-only variant without ROI worsened HD95 because reconstruction was not guided by localized foreground normalization, making it sensitive to mask position, scale variation, and background-dominated regions.

4.4. Comparative Evaluation with Existing Segmentation Methods

The proposed method was compared with Attention U-Net, U-Net, U-Net++, DeepLabV3+, and YOLOv8n-seg using the same dataset split, preprocessing, image size, augmentation, and testing subset, as shown in Table 7. All comparison models were trained using pretrained initialization and fine-tuned on the same training subset, with identical validation and testing splits, preprocessing, augmentation strategy, evaluation protocol, and comparable training budgets. Hyperparameters were fixed consistently across comparison models where applicable, and no extensive model-specific tuning was performed. The proposed method achieved the best DSC, IoU, HD95, and ASSD in this experimental setting, indicating better overlap quality, boundary stability, and anatomical consistency on the tested subset. Although Attention U-Net, U-Net, and U-Net++ achieved acceptable overlap results, their boundary errors remained higher in noisy ultrasound images, while DeepLabV3+ still underperformed YOLOv8n-seg and the proposed method. Overall, ROI-Based Shape-Prior Reconstruction further improves YOLOv8n-seg, especially in boundary consistency and morphology preservation.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 7. Comparative quantitative performance evaluation among segmentation methods

Method	DSC	IoU	HD95	ASSD
Attention U-Net	0.8656	0.8104	31.97	8.56
U-Net	0.8746	0.8200	18.54	5.68
U-Net++	0.8775	0.8256	24.94	8.10
DeepLabV3+	0.8897	0.8283	24.35	6.65
YOLOv8n-seg	0.9282	0.8671	15.06	5.38
Proposed Method	**0.9302**	**0.8708**	**14.18**	**5.20**

4.5. Qualitative Visualization Analysis

Qualitative visualization compared raw YOLOv8n-seg and reconstructed masks based on boundary alignment, anatomical consistency, and overlay results. Figures 10 and 11 show that YOLOv8n-seg localizes the cerebellum well but still produces unstable boundaries, while ROI-based shape-prior reconstruction yields smoother and more anatomically consistent masks, with clearer boundary improvements in moderate and difficult cases involving varying boundary clarity, contrast, artifacts, and deformation. Figure 12 further shows that Attention U-Net, U-Net, U-Net++, and DeepLabV3+ suffer from fragmentation, deformation, and boundary instability, whereas the proposed method produces smoother, more compact, and anatomically consistent fetal cerebellum segmentation. However, failures occurred when the initial YOLOv8n-seg mask was severely mislocalized, incomplete, or highly distorted, because the reconstruction depends on the ROI extracted from the raw mask; inaccurate localization may cause imperfect cropping and preserve or amplify the initial error. Thus, the module improves morphology when the initial ROI is reasonably localized but cannot fully recover severe localization failures. Another limitation is that the autoencoder learns the dominant training-mask shape distribution and may reconstruct a more average cerebellum shape, potentially smoothing or suppressing abnormal morphology or rare anatomical variations. Therefore, this method should be interpreted as a boundary-refinement approach, not evidence of clinical validity, especially without external validation and abnormal-case evaluation.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

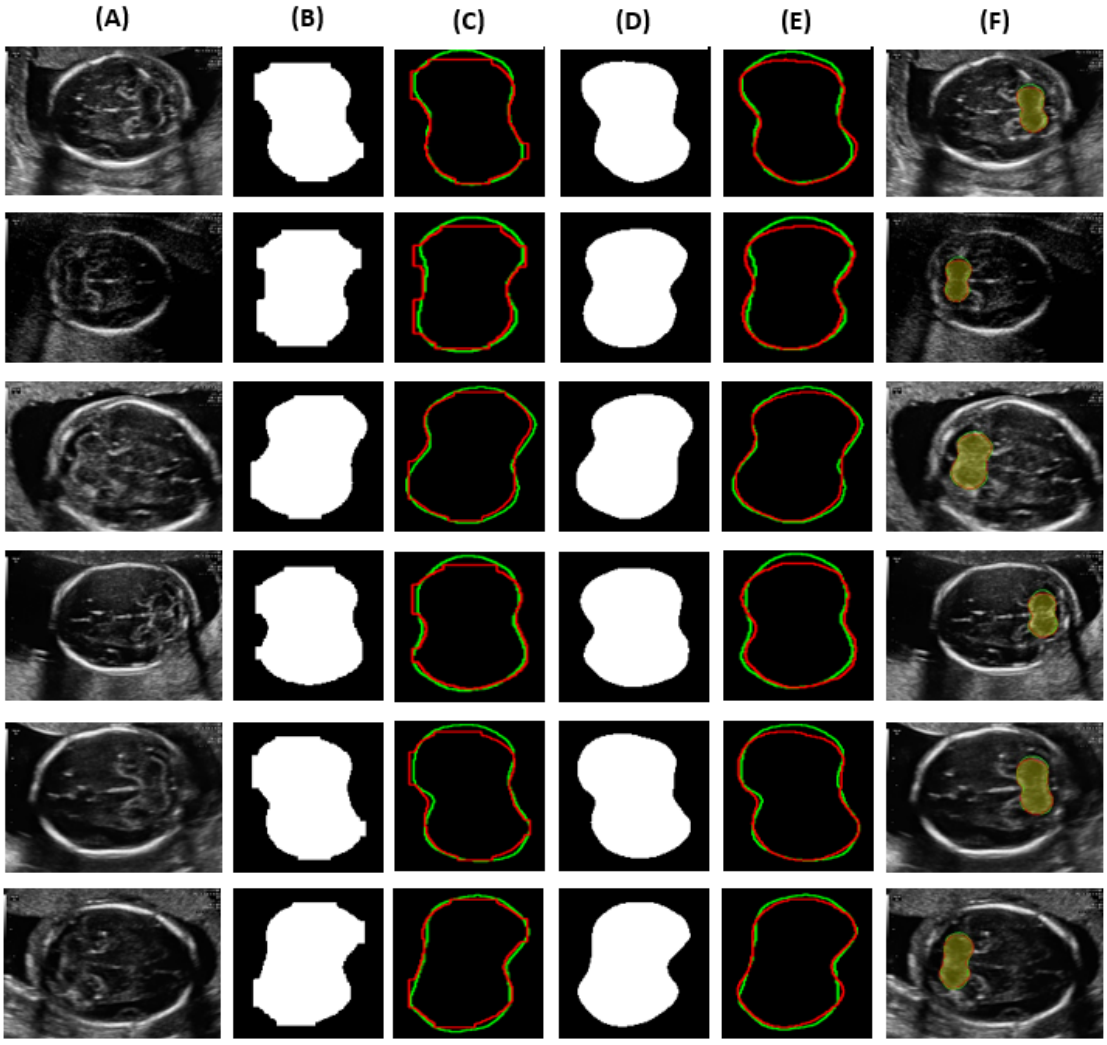

Figure 10. Qualitative comparison of fetal cerebellum ultrasound segmentation results. (A) Original ultrasound image, (B) raw YOLOv8n-seg mask, (C) overlay comparison between ground truth (green) and raw YOLOv8n-seg prediction (red), (D) reconstructed mask using ROI-Based Shape-Prior Reconstruction, (E) overlay comparison between ground truth (green) and reconstructed segmentation (red), (F) final reconstructed segmentation visualization on the original ultrasound image.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

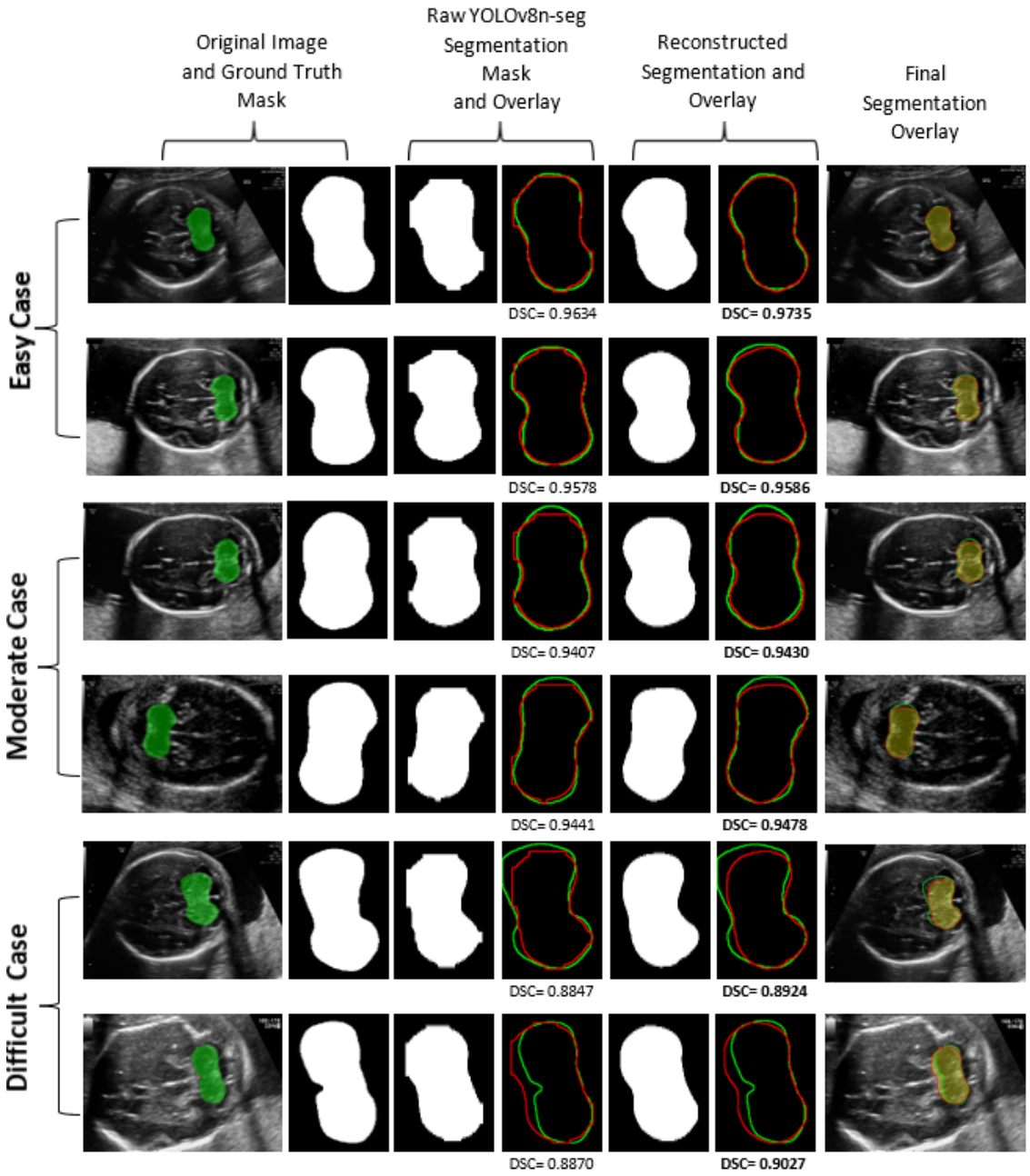

Figure 11. Qualitative comparison of fetal cerebellum segmentation under different difficulty levels: easy, moderate, and difficult cases. Each row shows the original ultrasound image, ground-truth mask, raw YOLOv8n-seg prediction, reconstructed segmentation result, and overlay comparison.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

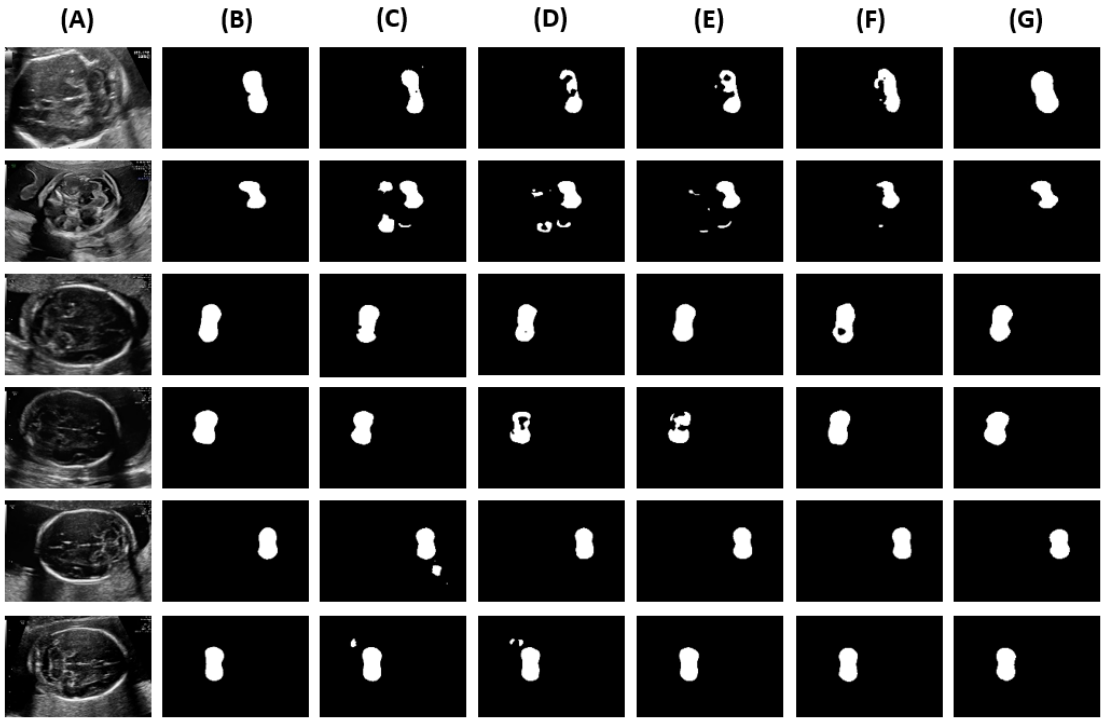

Figure 12. Qualitative comparison among segmentation methods on fetal cerebellum ultrasound images. (A) Original ultrasound image, (B) ground-truth mask, (C) Attention U-Net, (D) U-Net, (E) U-Net++, (F) DeepLabV3+, and (G) Proposed ROI-Based Shape-Prior Reconstruction.

4.6. Discussion

The results show that integrating YOLOv8n-seg with ROI-Based Shape-Prior Reconstruction improves fetal cerebellum segmentation quality. DSC increased from 0.9282 to 0.9302 and IoU from 0.8671 to 0.8708, although the gains were small and not statistically significant because the YOLOv8n-seg baseline was already high. The main improvement is clearer in boundary metrics, with HD95 decreasing from 15.06 to 14.18 and ASSD from 5.38 to 5.20. This indicates better boundary consistency, as the shape-prior autoencoder reduces local irregularities, protrusions, and fragmented regions by reconstructing ROI masks based on learned cerebellum shape distributions.

The high autoencoder validation performance should be interpreted carefully because it was trained and validated on normalized ROI binary masks, not full ultrasound images. Thus, its validation Dice mainly reflects reconstruction ability in ROI space, while final test performance depends on the full pipeline, including YOLOv8n-seg localization, raw mask quality, ROI extraction, reconstruction, and paste-back alignment. The

reconstruction stage increases inference time compared with YOLOv8n-seg alone, from 16.08 ms/image to 24.16 ms/image, corresponding to an additional processing cost of 8.09 ms/image. However, the module remains lightweight with approximately 0.77 million trainable parameters and operates only on a 128 × 128 cropped binary mask, limiting the added cost to ROI-level refinement.

Compared with conventional encoder-decoder methods, the proposed method benefits from strong YOLOv8n-seg localization and shape-prior reconstruction for improved morphological consistency. Unlike boundary-aware methods, it does not require explicit boundary annotation or an additional boundary supervision branch. Unlike shape-aware segmentation networks, it does not modify the main segmentation backbone or impose shape constraints during YOLOv8n-seg training. Compared with autoencoder-based refinement methods, its ROI-based design allows the autoencoder to focus on the cerebellum while reducing background influence.

However, several limitations remain. Shape-prior reconstruction may introduce over-smoothing, which can reduce fragmented masks and irregular contours but may also remove subtle anatomical boundary variations. Since the shape prior was learned from 294 images, it may not fully represent broader cerebellum variability across gestational age, scanner settings, fetal position, operator differences, or abnormal cases. The method also depends on the initial YOLOv8n-seg prediction; incorrect localization or severely distorted masks may cause ROI cropping errors that the autoencoder cannot fully recover.

The clinical relevance should also be interpreted cautiously. Although the proposed method improves segmentation quality and boundary consistency, this study did not directly evaluate biometric measurement accuracy, such as transverse cerebellar diameter. Therefore, the findings indicate improved segmentation performance rather than direct evidence of improved clinical measurement accuracy. Future work should include external validation, abnormal cerebellum cases, multi-operator annotation analysis, direct biometric measurement evaluation, and real-time deployment testing in clinical ultrasound workflows.

4.7. Limitations

This study has several limitations. The dataset is relatively small, consisting of 294 fetal ultrasound images from a single public source, without external validation. The study also focuses only on one anatomical object, namely the fetal cerebellum. In addition, annotation bias may exist because fetal cerebellum boundaries in ultrasound images are often ambiguous, and the reconstruction stage depends on the initial YOLOv8n-seg localization quality. Patient-level splitting also could not be performed because patient identifiers were unavailable, so potential patient-level overlap across subsets could not be fully excluded.

The shape-prior autoencoder may over-smooth reconstructed masks and may not fully preserve abnormal or atypical cerebellum morphology because it was learned from a limited dataset. This study also evaluated segmentation quality only, without directly assessing biometric measurement accuracy or real-time clinical deployment. Therefore, further studies are needed using larger multi-center datasets, abnormal cases, multi-operator annotations, biometric measurement evaluation, and real-time clinical testing.

5. CONCLUSION

This study proposed an ROI-Based Shape-Prior Reconstruction method to refine YOLOv8n-seg segmentation outputs for fetal cerebellum ultrasound images. Under the tested dataset and experimental protocol involving 294 fetal ultrasound images, the proposed method improved DSC from 0.9282 to 0.9302, IoU from 0.8671 to 0.8708, HD95 from 15.06 to 14.18, and ASSD from 5.38 to 5.20 compared with raw YOLOv8n-seg segmentation. These results indicate that the main benefit of the proposed method is more evident in boundary-based metrics than in overlap-based metrics, although the improvements were modest and not statistically significant. However, the findings should be interpreted within the scope of the evaluated dataset, since this study used a single public dataset and did not perform external validation. Therefore, clinical applicability cannot be concluded at this stage. Future work will focus on validation using larger multi-center datasets, cross-site evaluation, abnormal and difficult ultrasound cases, multi-operator annotation analysis, direct biometric measurement evaluation, robustness testing under poor image quality, and real-time inference evaluation.

REFERENCES

[1] Z. Sun, Y. Chen, and Q. Su, "Prenatal ultrasound for the diagnosis of the cerebellar abnormalities: a meta-analysis," *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 38, no. 1, Dec. 2025, doi: 10.1080/14767058.2025.2453997.

[2] N. Peñuelas *et al.*, "Gestational age assessment by ultrasound cerebellar measurements in fetal and perinatal deaths," *Am. J. Obstet. Gynecol.*, vol. 232, no. 6, pp. 559.e1-559.e10, Jun. 2025, doi: 10.1016/j.ajog.2024.11.016.

[3] Q. Wang, D. Zhao, H. Ma, and B. Liu, "Advanced fetal cerebellar vermis segmentation and gestational age prediction in ultrasound imaging for prenatal neural development assessment," *Eng. Appl. Artif. Intell.*, vol. 164, Art. no. 113315, Jan. 2026, doi: 10.1016/j.engappai.2025.113315.

[4] O. Rainio, E. Roshan, S. M. Hosseini, R. Rehman, J. Okenwa, and R. Klén, "Deep Learning for Medical Ultrasound Image Segmentation: A Systematic Review of the Current Research," *Journal of Imaging Informatics in Medicine*, Mar. 2026, doi: 10.1007/s10278-026-01931-1.

[5] L. Xiao, J. Song, X. Xie, and C. Fan, "Enhanced medical image segmentation using U-Net with residual connections and dual attention mechanism," *Eng. Appl. Artif. Intell.*, vol. 153, Aug. 2025, doi: 10.1016/j.engappai.2025.110794.

[6] V. Asadpour and F. Xie, "Artificial intelligence for medical imaging: a review of U-Net technology for anatomical feature analysis," *Intelligent Medicine*, Feb. 2025, doi: 10.1016/j.imed.2025.07.003.

[7] K. G. Khushubu *et al.*, "TransUNetB: An advanced Transformer–UNet framework for efficient and explainable brain tumor segmentation," *Inform. Med. Unlocked*, vol. 59, Jan. 2025, doi: 10.1016/j.imu.2025.101706.

[8] Z. Cai, K. Zhou, and Z. Liao, "A Systematic Review of YOLO-Based Object Detection in Medical Imaging: Advances, Challenges, and Future Directions," *Computers, Materials & Continua*, vol. 85, no. 2, pp. 2255–2303, 2025, doi: 10.32604/cmc.2025.067994.

[9] A. M. Mostafa *et al.*, "Optimized YOLOv8 for enhanced breast tumor segmentation in ultrasound imaging," *Discover Oncology*, vol. 16, no. 1, Dec. 2025, doi: 10.1007/s12672-025-02889-2.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[10] J. Chen *et al.*, "A deep learning-based multimodal medical imaging model for breast cancer screening," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-99535-2.

[11] H. Qiu, C. Zhong, C. Gao, and C. Huang, "Boundary-enhanced local-global collaborative network for medical image segmentation," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-93875-9.

[12] B. Li, W. Zhou, and H. Li, "A hybrid CNN-Transformer network integrating multiscale spatially detailed features for medical image segmentation," *PLoS One*, vol. 21, no. 4, Art. no. e0345549, Apr. 2026, doi: 10.1371/journal.pone.0345549.

[13] Y. Zhou *et al.*, "Efficient few-shot medical image segmentation via self-supervised variational autoencoder," *Med. Image Anal.*, vol. 104, Aug. 2025, doi: 10.1016/j.media.2025.103637.

[14] X. Yu, L. Teng, D. Zhang, J. Zheng, and H. Chen, "Attention correction feature and boundary constraint knowledge distillation for efficient 3D medical image segmentation," *Expert Syst. Appl.*, vol. 262, Mar. 2025, doi: 10.1016/j.eswa.2024.125670.

[15] P. Zhang, Y. Cheng, and S. Tamura, "Shape prior-constrained deep learning network for medical image segmentation," *Comput. Biol. Med.*, vol. 180, Sep. 2024, doi: 10.1016/j.compbiomed.2024.108932.

[16] A. Vatanparast, M. Fateh, H. Mashayekhi, and S. Ferdowsi, "SS_CASE_UNet: an attention-enhanced semi-supervised framework for fetal cerebellum segmentation in ultrasound images," *Sci. Rep.*, vol. 15, no. 1, Art. no. 44536, Dec. 2025, doi: 10.1038/s41598-025-28201-4.

[17] T. Wang *et al.*, "DCCE-UNet: a difference and context-aware contrast enhanced framework for ultrasound image segmentation," *BMC Med. Imaging*, vol. 25, no. 1, Dec. 2025, doi: 10.1186/s12880-025-01954-0.

[18] X. Xiao *et al.*, "Deep Learning-Based Medical Ultrasound Image and Video Segmentation Methods: Overview, Frontiers, and Challenges," *Sensors*, vol. 25, no. 8, Apr. 2025, doi: 10.3390/s25082361.

[19] D. Dai, C. Dong, H. Huang, F. Liu, Z. Li, and S. Xu, "Improving the performance of medical image segmentation with instructive feature learning," *Med. Image Anal.*, vol. 107, Jan. 2026, doi: 10.1016/j.media.2025.103818.

[20] Q. He *et al.*, "Masked pretraining of U-Net for ultrasound image segmentation," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-11688-2.

[21] F. Neha, D. Bhati, D. K. Shukla, S. M. Dalvi, N. Mantzou, and S. Shubbar, "An analytics-driven review of U-Net for medical image segmentation," *Healthcare Analytics*, vol. 8, p. 100416, Dec. 2025, doi: 10.1016/j.health.2025.100416.

[22] H. Liu, Y. Chen, R. Wang, M. Li, and Z. Li, "MFA-Deeplabv3+: an improved lightweight semantic segmentation algorithm based on Deeplabv3+," *Complex and Intelligent Systems*, vol. 11, no. 10, Oct. 2025, doi: 10.1007/s40747-025-02028-y.

[23] A. Garbaz, Y. Oukdach, S. Charfi, M. El Ansari, L. Koutti, and M. Salihoun, "GSAC-UFormer: Groupwise Self-Attention Convolutional Transformer-Based UNet for Medical Image Segmentation," *Cognit. Comput.*, vol. 17, no. 2, Apr. 2025, doi: 10.1007/s12559-025-10425-1.

[24] Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, "Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods," *Tomography*, vol. 11, no. 5, p. 52, Apr. 2025, doi: 10.3390/tomography11050052.

[25] R. Sapkota *et al.*, "YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series," *Artif. Intell. Rev.*, vol. 58, no. 9, Sep. 2025, doi: 10.1007/s10462-025-11253-3.

[26] C. Natarajan, S. Rajendran, M. S. Vinmathi, and R. M. Gomathi, "ROI-guided relational YOLO–SegNet transformer for lightweight bone tumor segmentation and classification from X-ray images," *Sci. Rep.*, vol. 16, no. 1, Art. no. 14603, Mar. 2026, doi: 10.1038/s41598-026-44297-8.

[27] L. Li, S. Lian, Z. Luo, B. Wang, and S. Li, "Contour-aware consistency for semi-supervised medical image segmentation," *Biomed. Signal Process. Control*, vol. 89, Mar. 2024, doi: 10.1016/j.bspc.2023.105694.

[28] A. L. Y. Hung *et al.*, "A Neural Conditional Random Field Model Using Deep Features and Learnable Functions for End-to-End MRI Prostate Zonal Segmentation," *Machine Learning for Biomedical Imaging*, vol. 3, no. August 2025, pp. 261–286, Aug. 2025, doi: 10.59275/j.melba.2025-gc4c.

[29] K. Dong *et al.*, "Position-aware representation learning with anatomical priors for enhanced pancreas tumor segmentation," *Neurocomputing*, vol. 616, Feb. 2025, doi: 10.1016/j.neucom.2024.128881.

[30] H. Abudukelimu *et al.*, "DVF-YOLO-Seg: A two-stage breast mass segmentation model with enhanced feature extraction and small lesion detection," *Digit. Health*, vol. 11, May 2025, doi: 10.1177/20552076251374192.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[31] H. Jebril, T. Pinetz, and H. Bogunović, "Shape Prior for Quality Assessment in OCTA via Denoising Autoencoders at the Segmentation Level," *IEEE Access*, vol. 13, pp. 187467–187476, 2025, doi: 10.1109/ACCESS.2025.3625745.

[32] F. A. Zaman, M. Jacob, A. Chang, K. Liu, M. Sonka, and X. Wu, "Latent diffusion for medical image segmentation: End-to-end learning for fast sampling and accuracy," *Biomed. Signal Process. Control*, vol. 114, Apr. 2026, doi: 10.1016/j.bspc.2025.109380.

[33] S. Gül, G. Cetinel, B. M. Aydin, D. Akgün, and R. Öztaş Kara, "YOLOSAMIC: A Hybrid Approach to Skin Cancer Segmentation with the Segment Anything Model and YOLOv8," *Diagnostics*, vol. 15, no. 4, Feb. 2025, doi: 10.3390/diagnostics15040479.

[34] M. Alzubaidi, M. Agus, M. Makhlouf, F. Anver, K. Alyafei, and M. Househ, "Large-scale annotation dataset for fetal head biometry in ultrasound images," *Data Brief*, vol. 51, Art. no. 109708, Dec. 2023, doi: 10.1016/j.dib.2023.109708.

[35] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Computerized Medical Imaging and Graphics*, vol. 95, p. 102026, Jan. 2022, doi: 10.1016/j.compmedimag.2021.102026.

[36] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Res. Notes*, vol. 15, Art. no. 210, Dec. 2022, doi: 10.1186/s13104-022-06096-y.

[37] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imaging*, vol. 15, no. 1, Aug. 2015, doi: 10.1186/s12880-015-0068-x.