



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/239825/>

Version: Accepted Version

Article:

Wu, D., Coldman, S., Zhou, G. et al. (2026) Automatic segmentation of placenta from MR images using a novel BiGC U-Net. IEEE Journal of Biomedical and Health Informatics, 30 (10). pp. 8813-8823. ISSN: 2168-2194

<https://doi.org/10.1109/jbhi.2026.3678395>

© 2026 The Authors. Except as otherwise noted, this author-accepted version of a journal article published in IEEE Journal of Biomedical and Health Informatics is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Automatic Segmentation of Placenta from MR images Using a Novel BiGC U-Net

Dengjiang Wu, Suzie Coldman, Guanglong Zhou, Xinshan Li, Elspeth Whitby, Lingzhong Guo, *Member, IEEE*

Abstract—Accurate segmentation of the placenta in Magnetic Resonance (MR) images is required for quantitative techniques such as texture and shape analysis, which have been proposed to improve placenta accreta spectrum (PAS) diagnostic rates. However, it is challenging due to the low contrast, image noise, blurred boundaries, and manual annotation variability. Hence, we proposed an enhanced U-Net architecture named BiGC U-Net, to automatically segment placental MR images. This deep learning (DL) structure incorporates a bidirectional gated convolutional module (BiGC) to capture complementary spatial dependencies, a hierarchical regularization mechanism (HRM) to enhance crosslayer semantic consistency, and an innovative data augmentation strategy to synthesize new images. The performance of BiGC U-Net was evaluated on three placental MR datasets: (1) public, (2) Sheffield Teaching Hospitals (STH) and (3) combined (public + STH + augmented), and compared against existing DL models including U-Net, Attention U-Net, ResNet, UNet++, TransUNet, nnUNet, and SSM-Mamba. The BiGC U-Net exhibited the best Dice similarity coefficient (90.74 ± 0.44), 95th percentile Hausdorff distance ($3.84 \text{ mm} \pm 0.53 \text{ mm}$) and relative volume difference (9.06 ± 0.41) in comparison with other DL models in the combined dataset. These findings indicate the effectiveness and robustness of the BiGC U-Net in accurately and automatically segmenting the placenta.

Index Terms—Deep learning, image segmentation, placental MR images, placenta, PAS, BiGC U-Net

I. INTRODUCTION

Placenta accreta spectrum (PAS) is caused by abnormal placental attachment to the uterine wall, potentially leading to massive hemorrhage, organ damage, and delivery complications [1]. Early and accurate PAS diagnosis is vital for ensuring adequate preoperative preparation and mitigating maternal and fetal health risks [2]. Magnetic resonance imaging (MRI) is an important diagnostic tool for PAS. However, due to the limited number of specialized radiologists available, there are wide regional variations in diagnostic accuracy [3]. Therefore, an urgent need exists to develop methods to assist clinicians in diagnosing PAS more accurately, by using automated methods to extract the placental region from MR images and

flag MR anomalies in the placenta.

Amongst various traditional deep learning (DL) algorithms which delineate anatomical regions of interest, U-Net [5] has shown excellent performance when applied to both MR and CT (Computed Tomography) images [6]. However, U-Net is limited to unidirectional feature extraction and struggles to model long-range dependencies [7]. Features at different levels in U-Net are learned independently, which weakens the semantic interaction between the encoder and decoder, leading to inefficient feature utilization preventing the model from combining global context with fine-grained structural details [8,9]. Consequently, more recent variants such as attention mechanism, UNet++, ResNet and Transformer model have been proposed to improve semantic fusion [8, 10-12]. For instance, in brain tumor applications, one study employed a deep learning ensemble framework for tumor segmentation in multimodal MRI, further utilizing it for feature extraction and survival prediction [40]. Meanwhile, YOLO-based approaches focus more on tumor detection, localization, and classification, thereby supporting subsequent analyses [41]. However, the problem of insufficient crosslayer communication remains unsolved [13].

Crosslayer communication is particularly challenging for placental MR images as the adjacent maternal tissues have a similar signal characteristic to the placenta. In addition, PAS can lead to blurred placental boundaries on MR images [14,15]. Recent studies have seen advancements in automatic placental segmentation using end-to-end 3D frameworks and tailored ensemble learning designs. These approaches addressed issues of positional and shape uncertainty alongside boundary ambiguity [6,16,36], with some incorporating prior knowledge of PAS-related MR features and radiomics for prenatal diagnosis and risk stratification [38,39].

Despite considerable progress in automatic placental segmentation, many challenges remain owing to anatomical variability, motion artifacts, and blurred boundaries, motivating researchers to explore more robust DL solutions [6,17,18]. In addition, placental image segmentation is limited by the scarcity of high-quality labeled datasets. The high inter-personal variability further limits the generalization of DL models, resulting in unstable performance [19,20]. These challenges underscore the need to develop network

This work was supported in part by the Yorkshire MedTech Translational Fund (YMTFPIAA027) (*Corresponding author: L. Guo*).

Dengjiang Wu, Guanglong Zhou, and Lingzhong Guo are from the School of Electrical and Electronic Engineering and Insigneo Institute, The University of Sheffield, Sheffield, UK (e-mail: dwu49@sheffield.ac.uk, gzhou9@sheffield.ac.uk, lguo@sheffield.ac.uk).

Suzie Coldman and Xinshan Li are from the School of Mechanical, Aerospace and Civil Engineering and Insigneo Institute, The University of Sheffield, Sheffield, UK (e-mail: scoldman1@sheffield.ac.uk, xinshan.li@sheffield.ac.uk).

Elsbeth Whitby is from the Sheffield Teaching Hospitals, Sheffield, UK (e-mail: e.whitby@sheffield.ac.uk).

architectures with more robust feature fusion mechanisms and greater contextual comprehension ability. Furthermore, developing effective data augmentation methods is also crucial for addressing limited data availability.

In this study, BiGC (bidirectional gated convolutional) U-Net, a novel DL architecture, was proposed to address some of these limitations. A bidirectional Mamba module [21] was incorporated to enhance feature extraction capability by capturing long-range dependencies and enabling rich contextual interactions in the spatial domain. Unlike conventional unidirectional encoders, the bidirectional structure enabled more comprehensive representation learning, which would be particularly beneficial in MR images with low contrast or blurred anatomical boundaries. Moreover, a hierarchical regularization mechanism (HRM) was designed to optimize crosslayer feature learning. By regulating semantic consistency and spatial refinement at different network depths, the proposed method would enhance feature fusion capacity and improve the information flow from global structures to local details. Furthermore, to mitigate the issue of limited training data, a data augmentation strategy was proposed. The strategy generated anatomically meaningful images and corresponding labels based on the mean shape of the placenta, enriching the training dataset.

The purpose of this study was therefore to investigate the effectiveness of the proposed BiGC U-Net model in segmenting the placenta from MR images, by comparing its performance with other popular DL models across three datasets: (1) public, (2) Sheffield Teaching Hospitals (STH) and (3) combined (public + STH + augmented).

II. METHODS

A. Data Augmentation

A data augmentation strategy was proposed using the deformable registration algorithm Sheffield Image Registration Toolkit (ShIRT) [22]. ShIRT was originally applied to musculoskeletal image registration [23], therefore modifications were made to apply it to placental MR images. The placenta data augmentation pipeline is shown in Fig. 1.

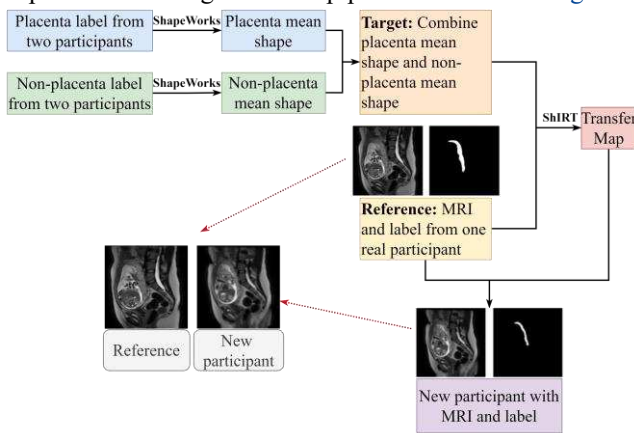


Fig. 1. Overview of ShIRT-based Data Augmentation Pipeline. The data augmentation process included anatomical labels fusion, ShIRT-based registration, and deformation map generation for synthesizing anatomically consistent augmented images.

First, each MR image was labelled with two distinctive regions: placenta and non-placenta. ShapeWorks [23] was used

to generate mean shapes [24] from two anatomically similar real participants for both placental and non-placental regions. Composite labels (known as the target) were then created by combining mean placental regions with mean non-placental regions. One of the real participant's labels was selected as the reference, and ShIRT was used to register the reference labels to the target to generate a deformation map. Finally, this deformation map was applied to the reference MR images and labels to generate a dataset of new images with labels as augmented participants. Compared with conventional geometric or intensity perturbation methods, this method introduced plausible anatomical structures, while maintaining the integrity of the placenta and surrounding tissues, thereby improving the robustness of the augmented images.

B. Overall Architecture of BiGC U-Net

In this study, an enhanced U-Net architecture was specifically developed for placental MR image segmentation. The framework retained the classic encoder-decoder structure of U-Net, while incorporating two novel modules: (1) a bidirectional gated convolutional (BiGC) module for advanced feature extraction and (2) a hierarchical regularization mechanism (HRM) to improve interlayer feature consistency.

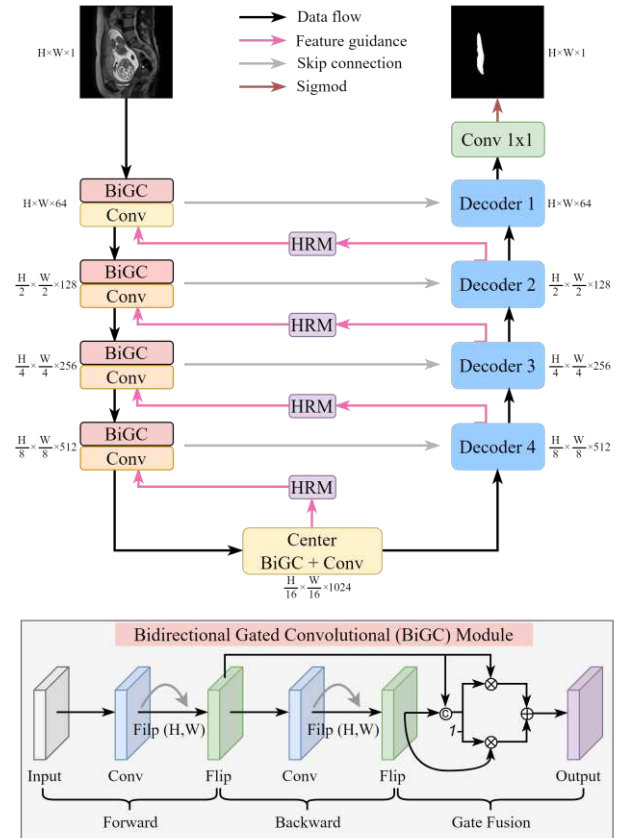


Fig. 2. Overview of the Enhanced U-Net Architecture. The figure illustrated an encoder-decoder framework, where the encoder stage was embedded with BiGC and HRM modules to connect multilevel features.

As shown in Fig. 2, the network featured a symmetric encoder-decoder structure. The encoder captured multiscale contextual information through stepwise down-sampling and convolutional blocks. The decoder progressively reconstructed the segmentation map through up-sampling operations and skip

connections, fusing the low-level spatial features and high-level semantic representations from the encoder. To enhance the feature-learning capability of the network, a BiGC module was integrated into each stage of the encoder to enable bidirectional spatial interaction and long-range dependency modeling. In addition, a HRM was inserted between the encoder and decoder paths to align the semantic features across multiple scales, thereby improving the global-to-local information flow and structural consistency. Furthermore, the integration of the BiGC and HRM enabled efficient end-to-end training with a reduced computational overhead.

C. Bidirectional Gated Convolutional (BiGC) Module

The BiGC module was proposed to enhance the contextual representation by extracting forward and backward spatial features. The design employed spatial flip operations to simulate bidirectional information flows and used a learnable gating mechanism to adaptively fuse the feature maps in both directions.

The input feature map $x \in \mathbb{R}^{C \times H \times W}$ was processed along two symmetric branches (Fig. 3). In the forward path, the input x was passed through a standard two-dimensional (2D) convolutional layer to extract forward features:

$$F_{forward} = \text{Conv}_{forward}(x) \quad (1)$$

In the reverse path, x was first flipped along the spatial dimensions (height and width), processed in the backward convolutional layers, and then flipped back to the original direction:

$$F_{backward} = \text{Flp}(\text{Conv}_{backward}(\text{Flp}(x))) \quad (2)$$

This “flip-convolution-flip” strategy allowed the network to perceive the reverse spatial information without introducing additional loop structures or positional coding, thus enriching the contextual representations with a minimal computational overhead.

To adaptively fuse forward and backward features, a lightweight gating mechanism was introduced. In this design, forward and inverse features were concatenated along the channel dimension and then fused in a 1×1 convolutional layer to generate a new representation. A sigmoid activation function was applied to generate gating weights:

$$G = \sigma(\text{Conv}_{1 \times 1}([F_{forward} \parallel F_{backward}])) \quad (3)$$

where $\sigma(\cdot)$ represents the sigmoid activation function, with the output value range of $[0,1]$. Moreover, $G \in [0,1]^{C \times H \times W}$ represents adaptive gating weights, and $\parallel$ indicates channel-wise concatenation. This gating mechanism allowed the model to achieve pixel-level control over the degree of dependence on forward and reverse features at each spatial location.

The final output was obtained via gated summation of forward and inverse features:

$$F_{out} = G \odot F_{forward} + (1 - G) \odot F_{backward} \quad (4)$$

The algorithm enabled the network to dynamically adapt to local structural and contextual variations while capturing complementary information from both directions, preserving fine anatomical details. Owing to the use of efficient convolutional operations and space flipping techniques, the entire module became lightweight and easy to integrate into a convolutional neural network.

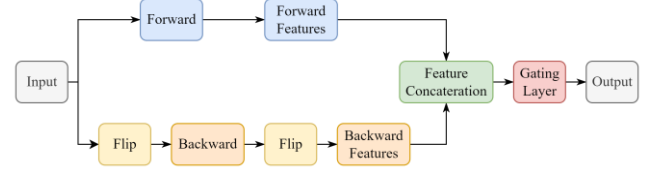


Fig. 3. Detailed Structure of the BiGC Module, which illustrates the forward and backward convolutional branches as well as the adaptive gating mechanisms.

D. Hierarchical Regularization Mechanism (HRM)

The HRM was proposed to enable feature learning in each layer to be guided by higher-level semantic representations. This mechanism was implemented based on the supervision of reference features, where the deeper reference features served as soft supervisors of the shallower ones. This approach improved crosslayer semantic consistency and facilitated a more coherent flow of information in the network.

As shown in Fig. 4, the module used the current feature map $x \in \mathbb{R}^{C \times H \times W}$ as well as the reference feature map $r \in \mathbb{R}^{C_r \times H_r \times W_r}$ extracted from a deep layer. To ensure spatial compatibility, the reference feature map was resized via bilinear interpolation to match the spatial resolution of the current feature map. The projected features were then computed as follows:

$$x^{proj} = \text{Conv}_{1 \times 1}(x) \quad (5)$$

$$r^{proj} = \text{Conv}_{1 \times 1}(r). \quad (6)$$

During training, the regularization loss was computed as the pixel-level mean square error (MSE) [25] between the projected features to ensure similarity between the current and reference projections:

$$\mathcal{L}_{reg} = \frac{1}{C \times H \times W} \sum_{i=1}^C \sum_{j=1}^H \sum_{k=1}^W (x_{i,j,k}^{proj} - r_{i,j,k}^{proj})^2 \quad (7)$$

where C represents the number of channels, H represents height, and W represents width. To integrate semantic bootstrapping into forward propagation, lightweight feature fusion was performed by linearly combining current and reference features:

$$x_{out} = x + \alpha \cdot r^{proj} \quad (8)$$

where $\alpha \in \mathbb{R}$ refers to a scalar weight that controls the influence of the reference features. This fusion strategy improved the characterization and regularization of the network while reducing its computational overhead.

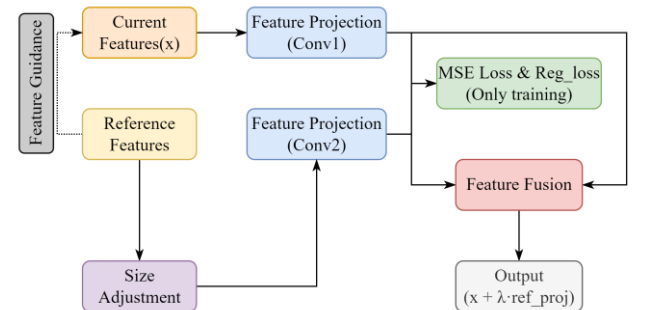


Fig. 4. Schematic of HRM Block. The structure projected and aligned deeper reference features to provide crosslayer guidance, thereby optimizing the learning process of the current features.

E. Loss Function Design

To optimize segmentation performance and ensure semantic consistency between feature layers, a composite loss function comprising two parts was defined: segmentation loss and hierarchical regularization loss.

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \lambda \cdot \mathcal{L}_{reg} \quad (9)$$

where $\mathcal{L}_{seg}$ and $\mathcal{L}_{reg}$ represent the segmentation and regularization losses, respectively. λ represents the weighting factor used to balance the contributions of the two terms.

The segmentation loss $\mathcal{L}_{seg}$ combined binary cross entropy (BCE) and the Dice losses to simultaneously penalize pixel-level classification errors and region-level mismatches.

$$\mathcal{L}_{seg} = \mathcal{L}_{BCE} + \mathcal{L}_{Dice} \quad (10)$$

where

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (11)$$

$$\mathcal{L}_{Dice} = 1 - \frac{2 \sum_{i=1}^N y_i \hat{y}_i + \epsilon}{\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{y}_i + \epsilon} \quad (12)$$

where y_i and $\hat{y}_i$ represent the true and predicted probabilities of the pixel i , respectively, and ϵ represents a small constant used for numerical stability.

III. MATERIALS AND EXPERIMENTS

A. Datasets

Three datasets of MR images were used in this study:

Public

This open-source dataset included 3,411 sagittal MR images from 131 PAS patients [26], acquired with a standard balanced turbo-field-echo sequence and a resolution of 512×512 pixels. Patient metadata and PAS subtypes were not available.

Sheffield Teaching Hospitals (STH)

This anonymized dataset included 3,135 sagittal MR images from 111 participants (50 healthy, 10 accreta, 8 increta, 27 percreta, 12 FGR, and 4 SGA) from the Sheffield Teaching Hospitals (STH), collected between September 2021 and July 2022. MR image protocol included a T2 single-shot fast spin echo with a repetition time (TR) of 1000ms and an echo time (TE) of 90ms on a 1.5T Siemens Advent or Siemens Aera scanner (Siemens Healthcare). The in-plane image resolution ranged from 192×192 to 512×512 pixels and slice spacing ranged from 0.49 to 1.82 mm.

Approval was obtained by the Central London Research Ethics Committee (IRAS number 260951) and the University of Sheffield's Research Ethics Committee (065440). PAS and its subtypes were determined through histology. FGR was determined in line with published guidance in Gordijn et al. [27] and/or a birthweight centile < 0.3 [28]. SGA was defined as a fetus with estimated fetal and/or birth weight centile < 10 th, with no FGR [29].

Three operators manually segmented the placental region in the sagittal plane, demonstrating a Dice coefficient of > 0.89 in an intra- and interobserver segmentation study [42]. Segmentation was reviewed by a Consultant Radiologist with > 20 years of experience in placental-fetal MR image diagnosis. To ensure data consistency and comparability for

DL, all images were resampled to 192×192 pixels with a slice thickness of 1.5 mm.

Augmented

A total of 16 augmented participants (sagittal images and placental labels) were generated from the STH dataset using the ShIRT-based augmentation method described in [Section II.A](#) (7 healthy, 3 accreta, 3 increta and 3 percreta). Images from healthy participants were used to generate augmented healthy participants, and similarly for each PAS subtypes. Augmented images were manually reviewed to ensure anatomical plausibility.

B. Implementation and Training Details

The proposed model was implemented in PyTorch 2.0.0 and trained on the Stanage high performance computing cluster equipped with an NVIDIA A100 GPU (80 GB memory). The training process used the Adam optimizer with an initial learning rate of $1e-4$. The default hyperparameters were $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1e-8$, and the batch size was 8 over a total of 100 epochs. After 8 epochs of linear warm-up, the learning rate decayed to $1e-6$ using a cosine annealing schedule. The loss function was defined as a combination of BCE (Binary Cross-Entropy) loss [30], Dice loss [31], and the hierarchical regularization loss to balance pixel-level accuracy and region-level overlap. The model employed a hierarchical regularization term without utilizing any additional explicit regularization (such as weight decay, L2 regularization, or dropout) in the optimizer or training settings. Model checkpoints were saved after each epoch, and the model version with the highest validation accuracy was selected for final testing.

The following number of participants were used for training, validation and testing (with an 8:1:1 split) in each dataset:

- (1) **Public Dataset:** 105 training, 14 validation and 12 testing (all PAS, no subtype information).
- (2) **STH Dataset:** 90 training, 11 validation and 10 testing (4 healthy, 2 accreta, 2 increta, and 2 percreta).
- (3) **Combined Dataset:** 206 training, 26 validation and 26 testing. The testing dataset included 12 public (PAS), 10 STH (4 healthy, 2 accreta, 2 increta and 2 percreta) and 4 augmented (1 healthy, 1 accreta, 1 increta and 1 percreta).

This partition was performed at the individual level to ensure that images used in the test dataset was independent from training and validation. During training, all input images were normalized to an intensity range of $[0,1]$. Standard data augmentation strategies were applied, including random flipping, rotation, and elastic deformation. To increase the reliability of the results, each run was repeated ten times with different random seeds, and the average performance of all the evaluation indicators was reported.

C. Evaluation Indicators

To evaluate the performance of the proposed segmentation framework, three metrics were used: Dice similarity coefficient [32], Hausdorff distance (HD) [33], and relative volume difference (RVD) [34].

Dice [32] measured the overlap between the predicted segmentation $|P|$ and the ground truth $|G|$, defined as:

$$Dice = \frac{2|P \cap G|}{|P| + |G|}, \quad (13)$$

where $|P|$ and $|G|$ represent the number of elements in the predicted segmentation mask and the true segmentation mask, respectively. $|P \cap G|$ represents the number of overlapping pixels between the predicted segmentation region and the true segmentation region.

HD [33] was used to assess the maximum boundary discrepancy between the predicted segmentation result P and the true segmentation result G . To minimize the impact of outliers, the 95th percentile HD (95HD) was used [35], which measured the distance where 95% of the predicted boundary points were within a threshold distance from the actual boundary points, defined as follows:

$$95HD(P, G) = \text{percentile}_{95} \left\{ \max \left(\sup_{p \in P} \inf_{g \in G} d(p, g), \sup_{g \in G} \inf_{p \in P} d(g, p) \right) \right\} \quad (14)$$

where $d(g, p)$ represents the Euclidean distance. The operator “sup” represents the supremum (maximum) over a set, indicating the largest distance, and “inf” represents the infimum (minimum), corresponding to the shortest distance to the nearest point. The 95th percentile was computed over all point-to-set distances, thereby excluding extreme outliers.

RVD [34] quantified the relative difference between the predicted and actual segmented volumes, as a measure of over-segmentation or under-segmentation. It was defined as:

$$RVD = \frac{|P - G|}{|P|} \times 100\%, \quad (15)$$

where P and G are the volumes (or numbers of pixels) of the predicted and the true segmentation results, respectively. In a 2D segmentation task, “volume” corresponds to the total number of pixels (i.e., area) of the segmented region. Lower RVD values indicate higher volumetric (area) accuracy.

The Shapiro–Wilk test [37] was conducted to examine whether the paired values followed a normal distribution. The results indicated that the differences between paired values in this study were non-parametric. Therefore, the Wilcoxon signed-rank test [23] was conducted to compare the performance of different DL methods, with $p < 0.05$ indicating statistical significance and $p < 0.01$ indicating high statistical significance.

IV. RESULTS AND DISCUSSION

A. Comparison with Existing Mainstream DL Methods

The proposed BiGC U-Net model was compared with U-Net [5], Attention U-Net [11], ResNet [12], UNet++ [8], nnUNet [6], TransUNet [10], and SSM-Mamba [43]. All models were trained under the same conditions and evaluated using Dice, 95HD, and RVD.

(1) Public Dataset

Table I and Fig. 5 show the evaluation metrics of the public dataset. BiGC U-Net achieved a mean Dice of 91.57%, 95HD of 7.88 mm, and RVD of 9.46%, significantly outperforming all other models ($p < 0.05$). Compared with nnUNet, Dice improved by 0.94% and 95HD decreased by 1.21 mm.

TABLE I

COMPARISON OF BiGC U-NET WITH MAINSTREAM DL METHODS ON PUBLIC DATASET.

Model	Dice (%)	95HD (mm)	RVD (%)	Parameters (M)	FLOPs (G)	Inference time (ms)
U-Net	88.45 ± 0.22	12.21 ± 0.43	14.09 ± 2.90	31.04	27.09	8.30
Attention U-Net	89.46 ± 0.35	10.47 ± 0.72	12.91 ± 2.22	31.39	27.70	11.60
ResNet	89.17 ± 0.41	8.95 ± 0.46	11.45 ± 1.57	32.44	28.62	11.19
UNet++	90.08 ± 0.85	9.65 ± 0.83	12.67 ± 1.16	11.18	16.77	13.34
TransUNet	88.86 ± 0.64	10.23 ± 0.68	12.56 ± 1.36	93.9	12.05	5.03
SSM-Mamba	90.10 ± 0.65	8.34 ± 0.60	11.22 ± 0.99	31.48	27.88	20.58
nnUNet	90.63 ± 0.31	9.09 ± 0.34	11.64 ± 0.89	8.12	7.17	8.98
BiGC U-Net	91.57 ± 0.27	7.88 ± 0.35	9.46 ± 0.15	39.40	34.37	14.82

(2) STH Dataset

Table II and Fig. 6 show the evaluation metrics of the STH dataset. BiGC U-Net achieved a Dice of 90.11%, with 95HD of 4.59mm and RVD of 9.71%. All metrics were significantly better than the other DL models tested ($p < 0.05$). Compared with nnUNet, 95HD decreased by 1.95mm, highlighting the superior boundary localization capabilities of the BiGC and HRM modules.

TABLE II

COMPARISON OF BiGC U-NET WITH MAINSTREAM DL METHODS ON STH DATASET.

Model	Dice (%)	95HD (mm)	RVD (%)	Parameters (M)	FLOPs (G)	Inference time (ms)
U-Net	87.77 ± 0.57	7.32 ± 1.42	12.95 ± 1.79	31.04	27.09	8.51
Attention U-Net	88.07 ± 0.48	6.27 ± 0.40	12.61 ± 1.05	31.39	27.70	11.87
ResNet	88.49 ± 0.38	5.64 ± 0.35	11.56 ± 0.36	32.44	28.62	12.40
UNet++	88.55 ± 0.20	5.77 ± 0.35	11.96 ± 0.92	11.18	16.77	13.29
TransUNet	86.77 ± 0.61	8.44 ± 1.06	12.63 ± 0.98	93.9	12.05	4.25
SSM-Mamba	87.71 ± 0.98	5.29 ± 0.45	12.65 ± 2.34	31.48	27.88	20.29
nnUNet	88.81 ± 0.37	6.53 ± 0.91	12.74 ± 0.81	8.12	7.17	8.88
BiGC U-Net	90.11 ± 0.53	4.59 ± 0.61	9.71 ± 0.50	39.40	34.37	13.91

(3) Combined Dataset

To evaluate the generalization capability, all DL models were also trained and tested on the combined dataset. The BiGC U-Net achieved a Dice of 90.74%, 95HD of 3.84 mm, and RVD of 9.06% (Table III and Fig. 7). BiGC U-Net consistently demonstrated superior performance in terms of segmentation accuracy, boundary and volumetric consistency against other DL models ($p < 0.05$), showing its strong generalizability across a heterogeneous cohort containing datasets from two centres.

TABLE III
COMPARISON BETWEEN BiGC U-NET AND MAINSTREAM DL METHODS ON COMBINED DATASET.

Model	Dice (%)	95HD (mm)	RVD (%)	Parameters (M)	FLOPs (G)	Inference time (ms)
U-Net	87.59 ± 0.26	6.29 ± 0.37	11.83 ± 0.64	31.04	27.09	10.11
Attention U-Net	88.32 ± 0.27	5.92 ± 0.45	11.20 ± 0.69	31.39	27.70	11.75
ResNet	88.30 ± 0.34	5.17 ± 0.13	10.49 ± 0.36	32.44	28.62	10.30
UNet++	88.77 ± 0.69	5.22 ± 0.30	10.31 ± 0.51	11.18	16.77	16.81
TransUNet	87.35 ± 0.91	14.86 ± 2.31	17.71 ± 1.32	93.9	12.05	7.82
SSM-Mamba	88.75 ± 1.24	4.23 ± 0.40	12.41 ± 1.92	31.48	27.88	18.87
nnUNet	89.75 ± 0.54	5.08 ± 0.21	10.26 ± 0.32	8.12	7.17	8.80
BiGC U-Net	90.74 ± 0.44	3.84 ± 0.53	9.06 ± 0.41	39.40	34.37	13.63

B. Ablation Study

A systematic ablation study was conducted by incrementally incorporating the BiGC module and HRM into the U-Net baseline model, to evaluate the independent contribution of each component. All models were trained under identical conditions and tested on three datasets using the same metrics.

Table IV presents the results of the ablation experiments performed on the public dataset. Incorporating the BiGC module increased Dice from 88.45% to 90.47%, decreased 95HD from 12.21 mm to 9.38 mm and decreased RVD from 14.09% to 11.74%. The addition of HRM alone also slightly improved performance (Dice: 89.95%, 95HD: 10.14 mm and RVD: 11.57%). The BiGC U-Net model achieved the highest performance (Dice: 91.57%, 95HD: 7.88 mm and RVD: 9.46%), demonstrating the combined complementary effects of both BiGC and HRM modules.

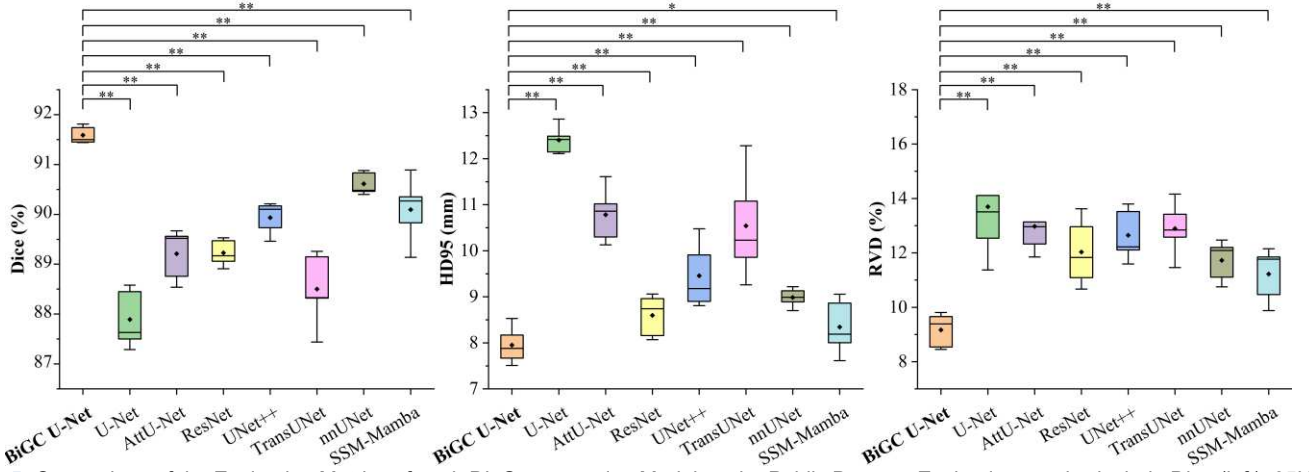


Fig. 5. Comparison of the Evaluation Metrics of each DL Segmentation Model on the Public Dataset. Evaluation metrics include Dice (left), 95HD (middle), and RVD (right). Significance level annotations: *p<0.05 and **p<0.01.

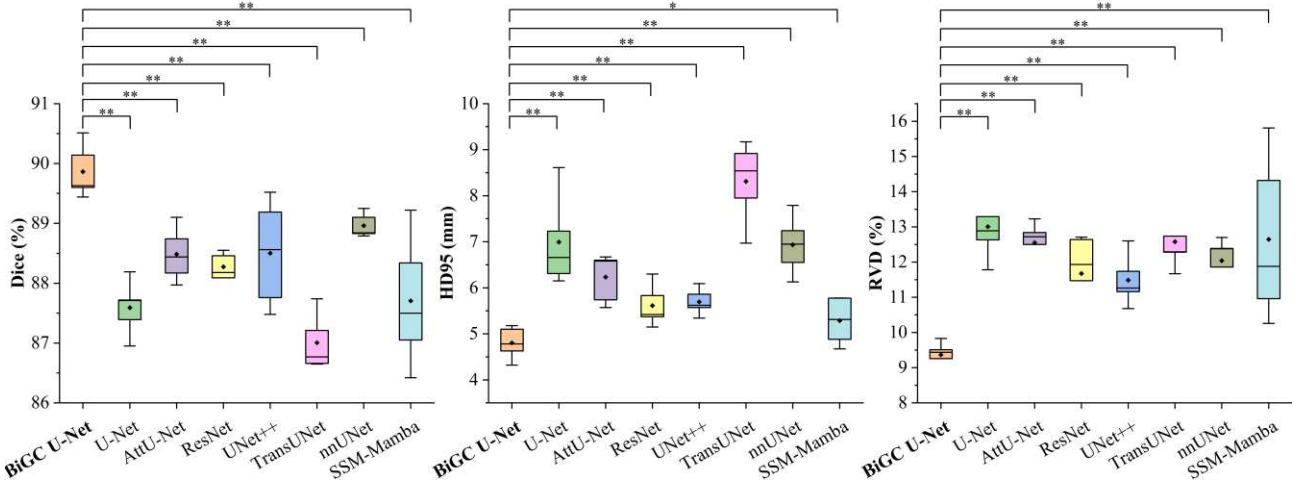


Fig. 6. Comparison of the Evaluation Metrics of each DL Segmentation Model on the STH Dataset. Evaluation metrics include Dice (left), 95HD (middle), and RVD (right). Significance level annotations: *p<0.05 and **p<0.01.

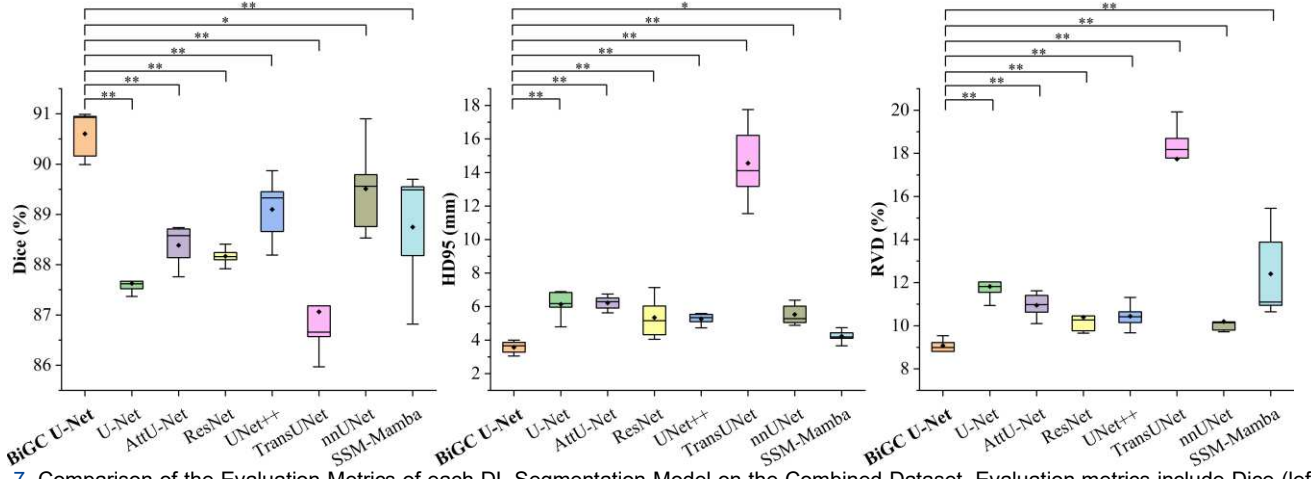


Fig. 7. Comparison of the Evaluation Metrics of each DL Segmentation Model on the Combined Dataset. Evaluation metrics include Dice (left), 95HD (middle), and RVD (right). Significance level annotations: * $p<0.05$ and ** $p<0.01$.

TABLE IV

RESULTS OF ABLATION STUDY ON PUBLIC DATASET.

Model	Dice (%)	95HD (mm)	RVD (%)
U-Net (Baseline)	88.45 $\pm$ 0.22	12.21 $\pm$ 0.43	14.09 $\pm$ 2.9
+ BiGC	90.47 $\pm$ 0.28	9.38 $\pm$ 0.97	11.74 $\pm$ 0.64
+ HRM	89.95 $\pm$ 0.43	10.14 $\pm$ 0.59	11.57 $\pm$ 1.22
BiGC U-Net	91.57 $\pm$ 0.27	7.88 $\pm$ 0.35	9.46 $\pm$ 0.15

Table V shows the results of the ablation study conducted on the STH dataset. The baseline U-Net achieved a mean Dice of 87.77%, 95HD of 7.32 mm and RVD of 12.95%. Incorporating the BiGC module and HRM separately saw improvements in all three metrics, with BiGC achieving a mean Dice of 89.71%, 95HD of 4.72 mm and RVD of 10.33%, and HRM achieving a mean Dice of 89.44%, 95HD of 4.97 mm, and RVD of 10.57%. The BiGC U-Net model delivered the best performance, achieving a mean Dice of 90.11%, 95HD of 4.59 mm and RVD of 9.71%.

TABLE V

RESULTS OF ABLATION STUDY ON STH DATASET.

Model	Dice (%)	95HD (mm)	RVD (%)
U-Net (Baseline)	87.77 $\pm$ 0.57	7.32 $\pm$ 1.42	12.95 $\pm$ 1.79
+ BiGC	89.71 $\pm$ 0.27	4.72 $\pm$ 0.40	10.33 $\pm$ 0.55
+ HRM	89.44 $\pm$ 0.86	4.97 $\pm$ 0.69	10.57 $\pm$ 0.62
BiGC U-Net	90.11 $\pm$ 0.53	4.59 $\pm$ 0.61	9.71 $\pm$ 0.50

Table VI shows the results on the combined dataset. The BiGC U-Net again achieved the highest mean Dice of 90.74%, the lowest mean 95HD of 3.84 mm and the lowest RVD of 9.06%, compared to introducing the BiGC and HRM modules separately.

TABLE VI

RESULTS OF ABLATION STUDY ON COMBINED DATASET.

Model	Dice (%)	95HD (mm)	RVD (%)
U-Net (Baseline)	87.59 $\pm$ 0.26	6.29 $\pm$ 0.37	11.83 $\pm$ 0.64
+ BiGC	90.12 $\pm$ 0.13	4.65 $\pm$ 0.21	9.69 $\pm$ 1.13
+ HRM	89.38 $\pm$ 0.15	5.62 $\pm$ 0.23	10.24 $\pm$ 0.81
BiGC U-Net	90.74 $\pm$ 0.44	3.84 $\pm$ 0.53	9.06 $\pm$ 0.41

The ablation study demonstrated that the BiGC and HRM modules played crucial roles in enhancing model performance.

The BiGC module effectively enhanced the capture of global and local features, thereby enhancing the accuracy of boundary delineation. The HRM stabilized the training process, reduced the sensitivity of the model to noise and outliers, and thereby helped suppress overfitting. The combination of both improved the overall segmentation accuracy, boundary consistency and enhanced the robustness of the model across different datasets. These results suggested that the BiGC U-Net is a better model for automatically segmenting placental MR images compared to existing DL models.

C. Sensitivity Analysis of the Loss Function

To investigate the impact of the hierarchical regularization weight (Eq. 7) on model performance, sensitivity analysis was conducted on three datasets using different weights (0.01, 0.05, 0.1, 0.2, and 0.5), as shown in Fig. 8. Such sensitivity analysis on weights has not been studied before [7].

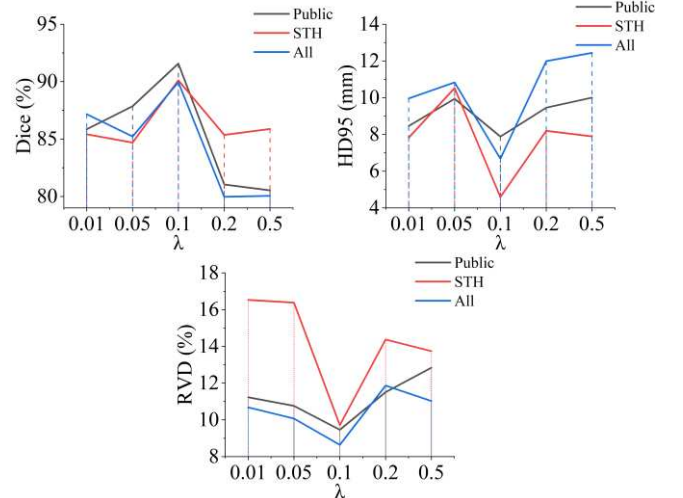


Fig. 8. Impact of Regularization Weight λ on Segmentation Performance.

Sensitivity analyses indicated that the performance of the model across the three datasets exhibited substantial variations with changes in λ . When λ was smaller (e.g., 0.01 or 0.05), the model exhibited insufficient semantic alignment between shallow and deep features, limiting its boundary delineation capability. When λ was larger (e.g., 0.2 or 0.5), the model

tended to overfit the projected reference features, resulting in performance degradation. The model achieved optimal performance at $\lambda = 0.1$. This value ensured crosslayer semantic consistency while avoiding overfitting.

D. Cross-dataset Generalization

To evaluate the robustness and generalization of the BiGC U-Net, cross-dataset analyses were conducted, where the model was trained on one dataset and tested on another. Two scenarios were considered:

Scenario 1. Public $\rightarrow$ STH dataset: Of the 131 PAS patients in the public dataset, 105 were used for training and 14 for validation. The model was then tested on 10 participants (4 healthy, 2 accreta, 2 increta and 2 percreta) from the STH dataset.

Scenario 2. STH $\rightarrow$ public dataset: Of the 111 participants in the STH dataset, 90 were used for training and 11 for validation. The model was then tested on 12 PAS patients from the public dataset.

Table VII shows the results of scenario 1. The BiGC U-Net achieved a mean Dice of 75.55%, 95HD of 17.04 mm and RVD of 19.39%, and significantly outperformed all other DL models ($p < 0.05$). Table VIII shows the results of scenario 2. Again, BiGC U-Net significantly outperformed all other models ($p < 0.05$), achieving a Dice of 71.43%, 95HD of 20.37 mm and RVD of 20.42%.

TABLE VII

CROSS-DATASET GENERALIZATION RESULTS FOR SCENARIO 1.

Model	Dice (%)	95HD (mm)	RVD (%)
U-Net	66.12 $\pm$ 2.32	23.87 $\pm$ 2.83	24.04 $\pm$ 1.22
Attention U-Net	62.91 $\pm$ 3.28	27.73 $\pm$ 2.87	24.61 $\pm$ 2.31
ResNet	67.74 $\pm$ 3.84	25.86 $\pm$ 0.29	25.27 $\pm$ 1.74
UNet++	64.25 $\pm$ 3.64	25.44 $\pm$ 1.94	23.38 $\pm$ 1.71
TransUNet	60.97 $\pm$ 2.42	29.72 $\pm$ 3.95	27.21 $\pm$ 1.24
SSM-Mamba	72.05 $\pm$ 1.56	21.71 $\pm$ 2.31	24.97 $\pm$ 1.75
nnUNet	66.86 $\pm$ 2.85	22.23 $\pm$ 1.41	24.74 $\pm$ 1.36
BiGC U-Net	75.55 $\pm$ 0.15	17.04 $\pm$ 0.56	19.39 $\pm$ 3.34

TABLE VIII

CROSS-DATASET GENERALIZATION RESULTS FOR SCENARIO 2.

Model	Dice (%)	95HD (mm)	RVD (%)
U-Net	62.39 $\pm$ 3.18	27.71 $\pm$ 1.67	28.51 $\pm$ 0.56
Attention U-Net	65.65 $\pm$ 3.69	25.78 $\pm$ 3.25	24.89 $\pm$ 1.07
ResNet	64.74 $\pm$ 4.35	25.44 $\pm$ 3.27	26.23 $\pm$ 4.21
UNet++	68.81 $\pm$ 2.25	22.02 $\pm$ 1.22	22.92 $\pm$ 2.39
TransUNet	54.95 $\pm$ 4.17	33.23 $\pm$ 2.26	27.23 $\pm$ 1.59
SSM-Mamba	63.48 $\pm$ 3.95	22.24 $\pm$ 2.53	21.97 $\pm$ 2.95
nnUNet	65.76 $\pm$ 2.97	22.83 $\pm$ 1.82	25.89 $\pm$ 1.56
BiGC U-Net	71.43 $\pm$ 0.41	20.37 $\pm$ 0.84	20.42 $\pm$ 0.90

These results demonstrated that our model possessed more transferable feature-learning capability compared to other mainstream DL models, maintaining a consistent level of segmentation performance across different datasets. However, compared with training and testing on the same dataset (Sec. A), the 95HD and RVD metrics in cross-dataset analyses

degraded, indicating a need for further improvements. Possible reasons for this degraded performance were: (1) limited size of datasets and no healthy participants in the public dataset (131 PAS cases only), (2) differences in MRI sequences, protocols and scanners across datasets (unknown for public dataset), leading to variations in intensity distribution and image resolution, and (3) variations in pre-processing and segmentation methods across centers (unknown for public dataset).

E. Qualitative Analysis and Visualization

Fig. 9 shows the augmented MR images based on methods described in Section II.A. Representative images of augmented healthy participants and different PAS subtypes are shown along with reference MR images and ground truth placental labels. The augmented MR images and labels aligned well with key maternal anatomical structures, providing an effective method to boost numbers in limited datasets.

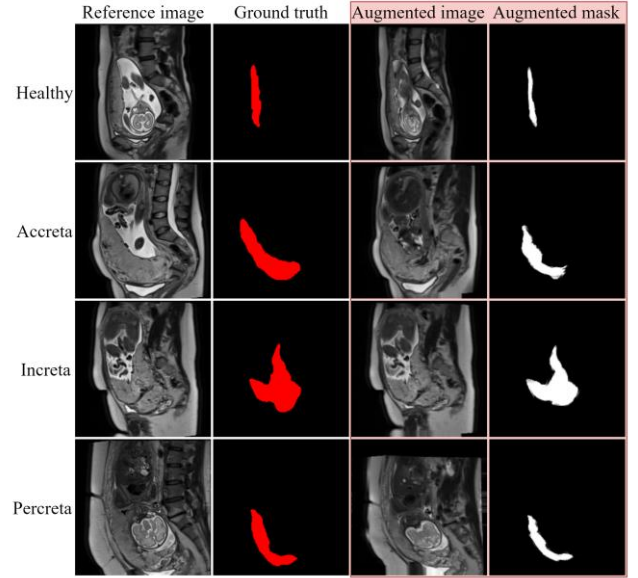


Fig. 9. Sagittal MR images of Healthy and PAS Placentas, showing reference MR image (column 1), reference placental label (column 2), augmented MR image (column 3) and its placental label (column 4).

Fig. 10 shows placental segmentation results from different data sources, comparing BiGC U-Net against mainstream DL models and ground truth manual segmentations. It is clear that the BiGC U-Net performed better than other algorithms in boundary delineation and capturing irregular placental shapes, especially in complex regions including: (a) areas with blurred boundaries between the placenta and uterine wall, (b) junction between placenta and amniotic fluid, and (c) irregular placental boundaries owing to lesion invasion, especially in increta and percreta cases where placental tissues penetrated the uterine muscle and sometimes adjacent organs. Overall, the BiGC U-Net better captured subtle placental structural variations while maintaining boundary consistency close to ground truth labels, achieving excellent overall segmentation accuracy and robustness when challenged with heterogeneous anatomy and large inter-personal variations.

F. Clinical Implications

The proposed BiGC U-Net model has been shown to achieve accurate and automated segmentation of placental MR images, while significantly improving the consistency and reliability of segmentation results compared to mainstream DL models. Furthermore, comparisons of the parameter count, FLOP count, inference time between BiGC U-Net and mainstream methods highlighted that the BiGC U-Net preserved segmentation accuracy whilst maintaining acceptable computational overhead and inference speed, supporting its feasibility for use in clinical research. Accurate placental segmentation holds significant value for evaluating placental morphology, texture and other metrics associated with placental dysfunction. This model, if implemented as part of an automated classification tool, could reduce the workload of manual annotation and contribute to enhanced efficiency in MR image analysis, and

the use of radiomics or DL based classification for early identification of placental dysfunction (e.g., PAS).

G. Limitations and Future Work

Although BiGC U-Net demonstrated strong performance and robustness in quantitative metrics and visualization results, certain limitations remain. First, the model performance degraded across different datasets, particularly for the 95HD and RVD, suggesting that its generalizability requires further enhancement. Second, the size of the datasets was modest and limited by the lack of healthy participants in the public dataset.

Future research could therefore incorporate domain adaptation and semi-supervised learning strategies to enhance cross-domain generalization capabilities, as well as expand the current datasets to include multi-center (three and more) and multi-scanner MR images for model training.



Fig. 10. Comparison of Placenta Segmentation Results from Different DL Models in representative cases from the STH (healthy, accreta, increta and percreta), augmented and public datasets. Different color boxes highlight typical errors in mainstream DL models. Blue boxes: false positive small fragments. Red boxes: gaps, holes and discontinuities. Green boxes: excessive segmentation of surrounding tissue.

V. CONCLUSION

In this study, an enhanced U-Net architecture, the BiGC U-Net was proposed, which integrated BiGC and HRM modules to improve the accuracy of placental MR image segmentation, demonstrating superior performance compared with mainstream DL models. Additionally, a ShIRT-based data augmentation strategy was introduced to generate augmented MR images and placental labels that were anatomically plausible, to increase the diversity of the dataset and number of images. Future work will focus on further expansion of the dataset using multi-centre data sources and the integration of the algorithm with other DL methods to automatically classify placental abnormalities in MR images.

REFERENCES

- [1] E. Jauniaux, J. C. Kingdom, and R. M. Silver, "A comparison of recent guidelines in the diagnosis and management of placenta accreta spectrum disorders," *Clin. Obstet. Gynaecol.*, vol. 72, pp. 102–116, Apr. 2021, doi: 10.1016/j.bpobgyn.2020.06.007.
- [2] E. Jauniaux, S. Collins, and G. J. Burton, "Placenta accreta spectrum: Pathophysiology and evidence-based anatomy for prenatal ultrasound imaging," *Am. J. Obstet. Gynecol.*, vol. 218, no. 1, pp. 75–87, Jan. 2018, doi: 10.1016/j.ajog.2017.05.067.
- [3] S. Hong, Y. Le, K. U. Lio, T. Zhang, Y. Zhang, and N. Zhang, "Performance comparison of ultrasonography and magnetic resonance imaging in their diagnostic accuracy of placenta accreta spectrum disorders: a systematic review and meta-analysis," *Insights into Imaging*, vol. 13, no. 1, Art. no. 50, 2022, doi: 10.1186/s13244-022-01192-w.
- [4] Y. Shen, N. Wu, J. Phang, J. Park, K. Liu, S. Tyagi, L. Heacock, S. G. Kim, L. Moy, K. Cho, K. J. Geras, "An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised

- localization,” *Med. Image. Anal.*, vol. 68, 2021, p. 101908. doi: 10.1016/j.media.2020.101908.
- [5] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” 2015, doi: 10.48550/arxiv.1505.04597.
- [6] J. Huang et al., “Deep-learning-based automatic segmentation of the placenta and uterine cavity on prenatal MR images,” in *Medical Imaging 2023: Computer-Aided Diagnosis* (Proc. SPIE, vol. 12465), Paper 124650N, 2023, doi: 10.1117/12.2653659.
- [7] C. Jiang, C. Zhu, H. Guo, G. Tan, C. Liu, and K. Li, “MCBL-UNet: A hybrid Mamba-CNN boundary enhanced light-weight UNet for placenta ultrasound image segmentation,” *IEEE J. Biomed. Health Inform.*, pp. 1–12, Aug. 2025, doi: 10.1109/jbhi.2025.3599716.
- [8] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A nested U-Net architecture for medical image segmentation,” 2018, doi: 10.48550/arxiv.1807.10165.
- [9] H. Cao et al., “Swin-UNet: Unet-like pure transformer for medical image segmentation,” 2021, doi: 10.48550/arxiv.2105.05537.
- [10] J. Chen et al., “TransUNet: Transformers make strong encoders for medical image segmentation,” 2021, doi: 10.48550/arxiv.2102.04306.
- [11] O. Oktay et al., “Attention U-Net: Learning where to look for the pancreas,” 2018, doi: 10.48550/arxiv.1804.03999.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.
- [13] J. Liu, Z. Hu, Y. Zhang, Y. Li, J. Yang, and H. Yu, “BiFormer attention - guided multiscale fusion Mask2former networks for fish abnormal behavior recognition and segmentation,” *Aquac. Res.*, vol. 2024, no. 1, p. 8892810, Jan. 2024, doi: 10.1155/are/8892810.
- [14] R. Aughwane, E. Ingram, E. D. Johnstone, L. J. Salomon, A. L. David, and A. Melbourne, “Placental MRI and its application to fetal intervention,” *Prenatal Diagnosis*, vol. 40, no. 1, pp. 38–48, 2020, doi: 10.1002/pd.5526.
- [15] Sitthipong Srisajjakul, Patcharin Prapaisilp, and Sirikan Bangchokdee, “Magnetic resonance imaging of placenta accreta spectrum: A step-by-step approach,” *Korean Journal of Radiology*, vol. 22, no. 2, pp. 198–212, 2021.
- [16] A. Jittou, K. E. Fazazy, and J. Riffi, “Placenta segmentation redefined: review of deep learning integration of magnetic resonance imaging and ultrasound imaging,” *Visual Computing for Industry, Biomedicine and Art*, vol. 8, no. 1, Art. no. 17, 2025, doi: 10.1186/s42492-025-00197-8.
- [17] M. A. Mazurowski, H. Dong, H. Gu, J. Yang, N. Konz, and Y. Zhang, “Segment anything model for medical image analysis: An experimental study,” 2023, doi: 10.48550/arxiv.2304.10517.
- [18] Z. Shen, P. Cao, J. Su, J. Yang, and O. R. Zaiane, “Adaptive mix for semi-supervised medical image segmentation,” 2024, doi: 10.48550/arxiv.2407.21586.
- [19] S. S. M. Salehi et al., “Real-time automatic fetal brain extraction in fetal MRI by deep learning,” in *Proc. IEEE Int. Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC, USA, 2018, pp. 720–724.
- [20] S. Mazdak Abulnaga et al., “Shape-aware segmentation of the placenta in BOLD fetal MRI time series,” 2023, doi: 10.48550/arxiv.2312.05148.
- [21] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, “Vision mamba: Efficient visual representation learning with bidirectional state space model,” 2024, doi: 10.48550/arxiv.2401.09417.
- [22] D. Barber and D. Hose, “Automatic segmentation of medical images using image registration: Diagnostic and simulation applications,” *J. Med. Eng. Technol.*, vol. 29, no. 2, pp. 53–63, Jan. 2005, doi: 10.1080/03091900412331289889.
- [23] Z. Lin, W. H. Henson, L. Dowling, J. Walsh, E. Dall’Ara, and L. Guo, “Automatic segmentation of skeletal muscles from MR images using modified U-Net and a novel data augmentation approach,” *Front. Bioeng. Biotechnol.*, vol. 12, p. 1355735, Feb. 2024, doi: 10.3389/fbioe.2024.1355735.
- [24] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models-their training and application,” *Comput. Vis. Image Underst.*, vol. 61, no. 1, pp. 38–59, Jan. 1995, doi: 10.1006/cviu.1995.1004.
- [25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, Massachusetts: The MIT Press, 2016.
- [26] T. Yang, “Placenta accreta spectrum disorders (PASDs),” Mendeley Data, V1, 2023, doi: 10.17632/284gwmf5bh.1.
- [27] S. J. Gordijn, I. M. Beune, B. T. Thilaganathan, A. Papageorgiou, A. A. Baschat, P. N. Baker, R. M. Silver, K. Wynia, W. Ganzevoort, “Consensus definition of fetal growth restriction: a Delphi procedure,” *UOG*, vol. 48, no. 3, pp. 333–339, 2016/09/01 2016, doi: <https://doi.org/10.1002/uog.15884>.
- [28] NHS England, *Saving Babies’ Lives: Version 3*, PRN00130, 2023. [Online]. Available: <https://www.england.nhs.uk/long-read/saving-babies-lives-version-3/#appendix-d-risk-assessment-surveillance-pathway-and-management-of-fgr>
- [29] N. Melamed, A. Baschat, Y. Yinon, A. Athanasiadis, F. Mecacci, F. Figueras, V. Berghella, A. Nazareth, M. Tahlak, H. D. McIntyre, F. D. Costa, “FIGO (International Federation of Gynecology and Obstetrics) initiative on fetal growth: best practice advice for screening, diagnosis, and management of fetal growth restriction,” *Int. J. Gynaecol. Obstet.*, vol. 152, no. S1, pp. 3–57, 2021, doi: 10.1002/ijgo.13522.
- [30] S. Jadon, “A survey of loss functions for semantic segmentation,” 2020, doi: 10.48550/arxiv.2006.14822.
- [31] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” in *Proc. Int. Conf. on 3D Vision (3DV)*, Stanford, CA, USA, 2016, pp. 565–571.
- [32] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945, doi: 10.2307/1932409.
- [33] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, “Comparing images using the Hausdorff distance,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 9, pp. 850–863, Aug. 1993, doi: 10.1109/34.232073.
- [34] R. V. Manjunath, Y. Gowda N, and H. M. Manu, “Automated liver segmentation from CT images using modified ResUNet,” *Gastroenterology & Endoscopy*, vol. 3, no. 2, pp. 93–104, 2025, doi: 10.1016/j.gande.2025.03.002.
- [35] A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool,” *BMC. Med. Imaging*, vol. 15, no. 1, p. 29, Aug. 2015, doi: 10.1186/s12880-015-0068-x.
- [36] I. Saito et al., “PlaNet-S: An automatic semantic segmentation model for placenta using U-Net and SegNeXt,” *Journal of Imaging Informatics in Medicine*, 2025, doi: 10.1007/s10278-025-01549-9.
- [37] B. W. Yap and C. H. Sim, “Comparisons of various types of normality tests,” *Journal of Statistical Computation and Simulation*, vol. 81, no. 12, pp. 2141–2155, 2011, doi: 10.1080/00949655.2010.520163.
- [38] H. Wang et al., “A deep learning pipeline using prior knowledge for automatic evaluation of placenta accreta spectrum disorders with MRI,” *Journal of Magnetic Resonance Imaging*, vol. 59, no. 2, pp. 483–493, 2024, doi: 10.1002/jmri.28770.
- [39] M. Danaei et al., “Machine learning applications in placenta accreta spectrum disorders,” *European Journal of Obstetrics & Gynecology and Reproductive Biology: X*, vol. 25, Art. no. 100362, 2025, doi: 10.1016/j.eurox.2024.100362.
- [40] D. Rastogi et al., “Deep learning-integrated MRI brain tumor analysis: feature extraction, segmentation, and Survival Prediction using Replicator and volumetric networks,” *Scientific Reports*, vol. 15, no. 1, Art. no. 1437, 2025, doi: 10.1038/s41598-024-84386-0.
- [41] M. F. Almufareh, M. Imran, A. Khan, M. Humayun, and M. Asim, “Automated brain tumor segmentation and classification in MRI using YOLO-Based deep learning,” *IEEE Access*, vol. 12, pp. 16189–16207, 2024, doi: 10.1109/ACCESS.2024.3359418.
- [42] S. Coldman, E. Whitby, and X. Li, “Segmentation of the placenta for the investigation of placenta accreta spectrum,” presented at the *16th BioMedEng Annual Conf. (BioMedEng24)*, London, U.K., Sep. 5–6, 2024.
- [43] A. Gu and T. Dao, “Mamba: Linear-time sequence modeling with selective state spaces,” 2023, doi: 10.48550/arxiv.2312.00752.