BDFR-Net: Boundary-Aware Dual-Domain Feature Reconstruction Network for Fetal Ultrasound Image Segmentation

Wenfeng Zhang[1 ⋆], Nana Liu[1 ⋆], Qibing Qin[2], Xin Huang[1], Cong Tan[1], Wei Hu[1 ✉], and Jieyun Bai[3 ✉]

[1] College of Computer and Information Science, Chongqing Normal University, Chongqing, China
wei369.workstation@cqnu.edu.cn
[2] School of Computer Engineering, Weifang University, Weifang, China
[3] College of Information Science and Technology, Jinan University, Guangzhou, China
baijieyun@jnu.edu.cn

Abstract. Accurate segmentation of the fetal head and pubic symphysis in ultrasound is critical for quantitative obstetric assessment and delivery decision-making. This task is severely hampered by inherent speckle noise and low signal-to-noise ratio, which obscure anatomical boundaries. Furthermore, semantic gaps in skip connections of existing networks degrade structural fidelity, leading to clinically significant segmentation errors. To overcome these limitations, we propose the Boundary-Aware Dual-Domain Feature Reconstruction Network (BDFR-Net), which synergistically models frequency-spatial complementarity to enhance boundary perception and structural fidelity. Specifically, the Dual-Domain Feature Decoupling and Boundary Enhancement (DDBE) is designed to tackle boundary ambiguity by decomposing features into frequency-decoupled edge responses and spatial contextual cues, which are then consolidated via channel-attentive cross-domain fusion. Furthermore, an Adaptive Context Alignment Module (ACAM) is embedded in skip connections to bridge semantic gaps. ACAM recalibrates cross-layer features through dual attention mechanisms, ensuring hierarchical consistency and mitigating feature drift. Extensive experiments on multiple fetal ultrasound datasets show that BDFR-Net achieves superior segmentation accuracy, attaining 86.83% mean Dice on the MICCAI IUGC 2024 benchmark and consistent improvements across cross-dataset evaluation, demonstrating strong generalization under distribution shift.

Keywords: Intrapartum Ultrasound · Medical Image Segmentation · Boundary Feature Enhancement.

⋆ W. Zhang and N. Liu contributed equally to this work.

1 Introduction

Accurate segmentation of the fetal head (FH) and pubic symphysis (PS) in transperineal ultrasound (TPU) is fundamental to quantitative intrapartum assessment [7]. These two anatomical structures serve as key landmarks for deriving obstetric biometrics such as the angle of progression, which informs clinical decisions regarding the mode of delivery [15]. Conventional assessment relies on manual palpation, which suffers from high inter-observer variability and limited reproducibility [14]. While TPU offers an objective and non-invasive alternative [22], as illustrated in Fig. 1, the inherent speckle noise, acoustic shadowing, and low signal-to-noise ratio of ultrasound imagery severely degrade tissue boundary contrast [21], making accurate automated delineation of FH and PS particularly difficult [29].

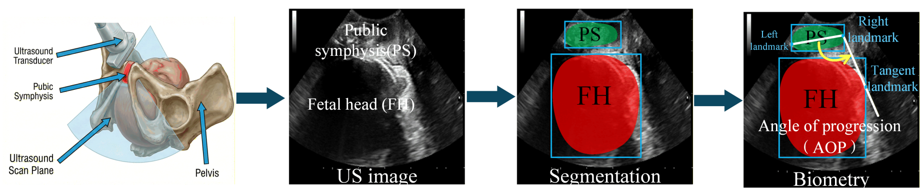

Fig. 1. Clinical pipeline for intrapartum ultrasound assessment. A transperineal ultrasound probe captures the fetal head (FH) and pubic symphysis (PS). Automated segmentation delineates both structures, from which anatomical landmarks are derived for quantitative obstetric measurements.

U-Net [20] and its variants have established the dominant paradigm for medical image segmentation. On the encoder side, CNN backbones [2, 4] and Transformer-based architectures [6, 12] have progressively enhanced feature abstraction, while decoder-side designs [3, 8, 19] improve boundary restoration during reconstruction. Meanwhile, skip-connection refinements such as nested dense pathways [30], full-scale aggregation [13], and channel-wise cross-attention [26] aim to bridge the semantic gap between encoder and decoder features. Despite these advances, existing methods predominantly rely on implicit feature transformation rather than explicit semantic alignment prior to fusion, leaving residual cross-level misalignment that compromises boundary continuity. Moreover, they operate exclusively in the spatial domain, where boundary responses and speckle-induced noise remain entangled in low-contrast ultrasound regions, limiting reliable fine boundary delineation.

Frequency-domain analysis provides a principled approach to decouple boundary and noise representations. The discrete wavelet transform (DWT) decomposes features into high-frequency subbands encoding directional edge transitions and low-frequency subbands preserving global structural continuity [16, 28], and has shown effectiveness in various reconstruction tasks [11]. However, existing frequency-based methods typically operate along a single decomposition direction or at a fixed scale [9, 10], failing to reconcile frequency-localized edge cues with spatially-distributed contextual dependencies. In ultrasound imaging where speckle artifacts simultaneously degrade both spectral and spatial representations, this isolated treatment fundamentally limits boundary delineation in anatomically ambiguous regions.

Motivated by these observations, we propose BDFR-Net, which addresses boundary ambiguity through synergistic frequency-spatial reconstruction while resolving cross-level semantic misalignment via explicit hierarchical alignment. The core contributions are: (1) a Dual-Domain Feature Decoupling and Boundary Enhancement (DDBE) module that disentangles frequency-domain spectral decompositions into directionally-refined edge responses and morphologically-stabilized structural priors, reconciling them with spatially-modeled long-range dependencies through cross-domain attentive aggregation to yield boundary-aware features preserving both edge fidelity and regional coherence. (2) an Adaptive Context Alignment Module (ACAM) that recalibrates cross-layer representations through dual spatial-channel attention before decoder fusion, explicitly mitigating distributional shift between encoder and decoder feature spaces. (3) Extensive experiments on multiple fetal ultrasound datasets demonstrate state-of-the-art segmentation accuracy and strong cross-dataset generalization.

2 Method

BDFR-Net follows an encoder-decoder paradigm tailored for boundary-ambiguous fetal ultrasound segmentation. As illustrated in Fig. 2, a PVTv2 [27] backbone pre-trained on ImageNet serves as the encoder, extracting hierarchical feature maps $\{E_i\}_{i=1}^4$ via spatial reduction self-attention. We report two variants, BDFR-B1 and BDFR-B2, built on the PVTv2-B1 and PVTv2-B2 encoders respectively. The decoder recovers anatomical structures through four cascaded stages, each comprising: (i) Channel Attention (CA) for suppressing task-irrelevant feature responses, (ii) Dual-Domain Feature Decoupling and Boundary Enhancement (DDBE) for frequency-spatial boundary enhancement, and (iii) Adaptive Context Alignment Module (ACAM) for cross-level semantic alignment prior to skip-connection fusion.

2.1 Dual-Domain Feature Decoupling and Boundary Enhancement

Spatial features entangle boundary responses with speckle artifacts, while single-scale frequency decomposition lacks long-range contextual modeling. DDBE reconciles both domains through three parallel pathways, preceded by a Channel Attention (CA) stage that suppresses speckle-dominated activations and highlights boundary-relevant features before spectral-spatial decoupling. Given $F \in \mathbb{R}^{C\times H\times W}$, CA processes global average-pooling $\mathcal{P}_{\text{avg}}$ and max-pooling $\mathcal{P}_{\max}$ through independent 1×1 bottleneck branches and fuses them through sigmoid

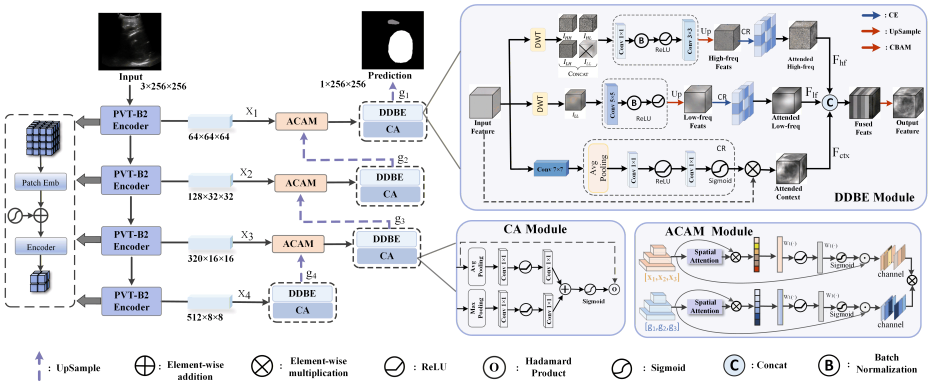

Fig. 2. Overview of the proposed BDFR-Net architecture. Left: the encoder-decoder pipeline with PVTv2 encoder stages and cascaded decoder blocks. Right: detailed structures of the Channel Attention (CA), Adaptive Context Alignment Module (ACAM), and Dual-Domain Feature Decoupling and Boundary Enhancement (DDBE) module.

gating to suppress speckle-correlated activations:

$$F_{\mathrm{CA}} = F \odot \sigma(\phi_1(\delta(\phi_1(\mathcal{P}_{\mathrm{avg}}(F)))) \oplus \phi_1(\delta(\phi_1(\mathcal{P}_{\max}(F))))) \tag{1}$$

where ϕ_k denotes a $k \times k$ convolution, $\oplus$ denotes element-wise addition and δ, σ denote ReLU and sigmoid respectively, ensuring the three subsequent pathways operate on boundary-salient features.

Since anatomical edges exhibit abrupt transitions along multiple orientations, the high-frequency branch decomposes F_{CA} via 2D-DWT with mother wavelet ψ into directional detail subbands over $\mathcal{D} = \{LH, HL, HH\}$:

$$\{S^{(d)}\}_{d \in \mathcal{D}} = \{\langle F_{\mathrm{CA}},\, \psi^{(d)}_{j,\mathbf{k}} \rangle\}_{d \in \mathcal{D}}, \quad S^{(d)} \in \mathbb{R}^{C \times \frac{H}{2} \times \frac{W}{2}} \tag{2}$$

These directional subbands are concatenated, projected, and refined to isolate true anatomical boundaries from isotropic speckle interference:

$$F_{\mathrm{hf}} = \mathrm{CR}\Big(\mathcal{U}\Big(\phi_3\Big(\delta\Big(\mathcal{B}\Big(\phi_1\Big(\bigoplus\nolimits_{d \in \mathcal{D}} S^{(d)}\Big)\Big)\Big)\Big)\Big)\Big) \tag{3}$$

where $\mathcal{B}$ denotes batch normalization, $\mathcal{U}$ denotes transposed convolution for spatial upsampling, and $\mathrm{CR}(\cdot)$ denotes the channel-recalibration operator defined in Eq. (5). The ϕ_1 projection first reduces the concatenated $3C$ channels to C, after which ϕ_3 refines directional edge responses in the compressed feature space.

Edge cues alone lack global morphological coherence. The low-frequency branch extracts the approximation subband $S^{(\ell)} = \langle F_{\mathrm{CA}},\, \varphi_{j,\mathbf{k}} \rangle \in \mathbb{R}^{C \times \frac{H}{2} \times \frac{W}{2}}$ via scaling function φ, and applies a 5×5 receptive field to preserve structural continuity:

$$F_{\mathrm{lf}} = \mathrm{CR}\Big(\mathcal{U}\Big(\delta\Big(\mathcal{B}\Big(\phi_5\Big(S^{(\ell)}\Big)\Big)\Big)\Big)\Big) \tag{4}$$

providing a morphological scaffold that maintains overall contour integrity for both FH and PS structures.

Since neither frequency branch captures long-range spatial dependencies, a spatial context branch employs 7×7 convolution with channel self-gating to model inter-regional coherence between spatially separated FH and PS boundaries:

$$F_{\text{ctx}} = F_{\text{CA}} \odot \text{CR}(\delta(\mathcal{B}(\phi_7(F_{\text{CA}})))), \quad \text{CR}(X) = \sigma(\phi_1(\delta(\phi_1(\mathcal{P}_{\text{avg}}(X))))) \tag{5}$$

where $\mathcal{P}_{\text{avg}}$ denotes global average pooling and the two-layer ϕ_1 bottleneck with sigmoid gating selectively amplifies structurally coherent activations via the Hadamard product $\odot$.

The three complementary representations are fused through sequential channel-spatial attention to suppress cross-domain interference:

$$F_{\text{DDBE}} = \mathcal{M}_{\text{sp}}\Big(\mathcal{M}_{\text{ch}}\Big(\phi_1\Big(F_{\text{hf}} \bigoplus F_{\text{lf}} \bigoplus F_{\text{ctx}}\Big)\Big)\Big) \tag{6}$$

where $\mathcal{M}_{\text{ch}}$ and $\mathcal{M}_{\text{sp}}$ denote the channel and spatial stages of CBAM, performing inter-branch importance weighting and region-specific saliency enhancement to produce boundary-discriminative features that preserve fine-grained edge continuity under speckle corruption.

2.2 Adaptive Context Alignment Module

The DDBE-enhanced decoder features are semantically rich but reside at a different abstraction level from the corresponding encoder features. Direct concatenation through skip connections ignores this representational asymmetry, leading to semantic drift that compromises boundary continuity [26]. ACAM addresses this by independently recalibrating encoder feature $g \in \mathbb{R}^{C_g \times H \times W}$ and DDBE-enhanced decoder feature $x \in \mathbb{R}^{C_x \times H \times W}$ through shared spatial-channel attention before their additive fusion.

For each stream $f \in \{g, x\}$, a 1×1 convolution projects f onto a single-channel response map, whose softmax-normalized values yield position-wise attention weights over spatial domain Ω of $H \times W$ positions:

$$\alpha_f(i) = \frac{\exp\left([\phi_1(f)]_i\right)}{\sum_{j \in \Omega} \exp\left([\phi_1(f)]_j\right)} \tag{7}$$

The weighted spatial aggregation then compresses each stream into a global context vector $\mathbf{c}_f \in \mathbb{R}^{C_f}$, where C_f denotes the channel dimension of f:

$$\mathbf{c}_f = \sum_{i \in \Omega} \alpha_f(i) \cdot f_i \tag{8}$$

To model inter-channel dependencies, the compressed context undergoes bottleneck refinement with layer normalization to produce stream-specific channel-wise modulation weights:

$$W_f = \sigma\left(\phi_1(\delta(\text{LN}(\phi_1(\mathbf{c}_f))))\right) \in \mathbb{R}^{C_f \times 1 \times 1} \tag{9}$$

where LN denotes layer normalization. Instantiating W_f for $f = g$ and $f = x$ yields W_g and W_x, which modulate and fuse the aligned feature streams:

$$F_{\mathrm{ACAM}} = W_g \odot g + W_x \odot x \tag{10}$$

The fused F_{ACAM} is upsampled, completing one decoder iteration; this repeats across all four levels.

2.3 Loss Function

Medical image segmentation suffers from severe class imbalance where foreground structures occupy small spatial regions. To address this, we employ a balanced combination of Dice loss and cross-entropy loss, where the former emphasizes region-level overlap sensitivity and the latter enforces pixel-level classification accuracy, jointly optimizing for both global structural consistency and local boundary precision.

3 Experiments

3.1 Datasets and Implementation

We evaluate BDFR-Net on three benchmarks. **FH-PS-AoP** [7] (MICCAI 2023 challenge) comprises two sub-datasets, **JNU-IFM** [17] (3,743 images, 51 patients) and **PSFHS** [5] (1,358 images, 1,124 subjects), providing 4,000 training and 700 test images annotated for FH and PS. As the official test set is unavailable, we split the 4,000 images 80%/20% into training and validation. To assess generalization under domain shift, we further train on JNU-IFM and test on PSFHS with no site or patient overlap. **MICCAI IUGC 2024** [1] provides 2,575 frames from 266 videos for training, evaluated on 300 independent standard-plane test images with no train-test overlap.

All experiments use PyTorch on a single NVIDIA RTX 4080 Super GPU with the AdamW optimizer (learning rate 1×10^{-4}, weight decay 1×10^{-4}). Images are resized to 256×256 with rotation, gray-scale, and gamma augmentation. The batch size is 16, trained for 100 epochs on all datasets except IUGC 2024 (50 epochs).

3.2 Comparison with State-of-the-Art Methods

We compare with eleven baselines under identical training configurations (Table 1). On FH-PS-AoP, BDFR-B2 attains 97.01% mean Dice and 2.668 HD, surpassing CFATransUnet (96.96%, 2.785). Under cross-dataset evaluation (JNU-IFM→PSFHS, no patient or site overlap), all baselines degrade sharply (CFATransUnet by 15.83 points), whereas BDFR-B2 retains 82.94% Dice and 1.434 ASD, exceeding TransUNet by 1.22% Dice and 23.9% ASD. On IUGC 2024 it reaches 86.83% Dice and 11.98 HD, a 12.4% HD reduction over CFATransUnet. Gains concentrate on the smaller, weaker-boundary PS, where BDFR-B2

Table 1. Quantitative comparison with state-of-the-art methods on FH-PS-AoP, Cross-dataset (JNU-IFM→PSFHS), and IUGC 2024 benchmarks. The best results are **bolded** and the second-best are underlined.

Dataset	Model	PS			FH			Mean (PSFH)		
		Dice (↑)	ASD (↓)	HD (↓)	Dice (↑)	ASD (↓)	HD (↓)	Dice (↑)	ASD (↓)	HD (↓)
FH-PS-AoP	U-Net [20]	0.9590	0.0648	2.5427	0.9790	0.0416	3.3334	0.9690	0.0532	2.9381
	AttU-Net [18]	0.9602	0.0577	2.3563	0.9790	0.0421	3.2564	0.9696	0.0499	2.8064
	ResU-Net [2]	0.9529	0.0778	2.7753	0.9761	0.0534	3.9331	0.9645	0.0656	3.3542
	FCT [23]	0.9599	0.0683	2.4754	0.9755	0.0427	3.3649	0.9677	0.0555	2.9202
	TransU-Net [6]	0.9525	0.0930	2.8233	0.9773	0.0423	3.4540	0.9649	0.0677	3.1387
	SwinU-Net [12]	0.9454	0.0969	3.0782	0.9743	0.0564	4.0825	0.9599	0.0767	3.5804
	U-Next [24]	0.9179	0.1285	4.2611	0.9662	0.0647	5.1146	0.9421	0.0966	4.6879
	SegNet [3]	0.9424	0.0951	3.3329	0.9690	0.0563	4.1825	0.9557	0.0757	3.7577
	G-CASCADE [19]	0.9433	0.0902	3.2017	0.9699	0.0558	4.0137	0.9566	0.0730	3.6077
	I2U-Net [8]	0.9488	0.0856	2.9390	0.9761	0.0461	3.5628	0.9625	0.0659	3.2509
	CFATransUnet [25]	0.9601	0.0560	2.3623	0.9791	0.0418	3.2072	0.9696	0.0489	2.7848
	BDFR-B1	0.9594	0.0603	2.4201	0.9794	0.0406	3.1269	0.9694	0.0504	2.7735
	BDFR-B2	**0.9606**	**0.0555**	**2.2981**	**0.9797**	**0.0391**	**3.0391**	**0.9701**	**0.0473**	**2.6680**
Cross-dataset	U-Net [20]	0.6732	4.0322	30.37	0.8901	1.4683	28.77	0.7817	2.7503	29.57
	AttU-Net [18]	0.6673	7.0841	37.19	0.8793	1.7833	26.57	0.7733	4.4337	31.88
	FCT [23]	0.6310	7.9114	22.86	0.8977	1.2799	22.03	0.7644	4.5957	22.45
	SegNet [3]	0.6396	6.4309	22.26	0.9075	1.0645	16.32	0.7736	3.7477	19.29
	SwinU-Net [12]	0.6666	3.7720	18.92	0.9052	1.2867	18.02	0.7859	2.5294	18.47
	R50U-Net [6]	0.7092	2.7611	23.88	0.8754	1.3259	29.81	0.7923	2.0435	26.85
	TransU-Net [6]	0.7255	2.5692	22.92	0.9089	1.2028	17.63	0.8172	1.8860	20.28
	U-Next [24]	0.5472	7.4457	32.15	0.8518	2.1225	30.10	0.6995	4.7841	31.13
	G-CASCADE [19]	0.6537	4.4623	17.19	0.8789	1.7490	19.08	0.7663	3.1057	18.14
	I2U-Net [8]	0.6537	4.3317	19.36	0.9077	0.9450	16.84	0.7807	2.6384	18.10
	CFATransUnet [25]	0.7308	2.9586	17.26	0.8918	1.6055	20.53	0.8113	2.2821	18.90
	BDFR-B1	0.7224	3.1032	**15.77**	**0.9247**	**0.6483**	**14.60**	0.8236	1.8757	**15.17**
	BDFR-B2	**0.7406**	**2.1151**	15.95	0.9183	0.7538	15.29	**0.8294**	**1.4344**	15.62
IUGC 2024	U-Net [20]	0.7613	1.6479	37.96	0.8953	1.1618	24.85	0.8283	1.4049	31.41
	AttU-Net [18]	0.7729	1.5462	39.35	0.9036	1.1784	20.53	0.8383	1.3623	29.94
	FCT [23]	0.7574	1.4261	14.16	0.9007	1.0296	18.59	0.8291	1.2279	16.38
	SegNet [3]	0.7560	2.2704	12.52	0.9171	0.8889	16.66	0.8366	1.5797	14.59
	SwinU-Net [12]	0.7543	1.5012	20.18	0.8939	1.0500	23.18	0.8241	1.2756	21.68
	R50U-Net [6]	0.7419	1.9013	21.05	0.9109	0.9021	19.50	0.8264	1.4017	20.28
	TransU-Net [6]	0.7699	1.6851	11.72	0.9223	0.6300	15.49	0.8461	1.1576	13.61
	U-Next [24]	0.6978	2.8567	23.28	0.8705	1.8775	24.92	0.7842	2.3671	24.10
	G-CASCADE [19]	0.7657	1.8644	14.43	0.9184	0.8462	16.21	0.8421	1.3553	15.32
	I2U-Net [8]	0.7698	1.2990	12.01	0.9087	0.8204	17.36	0.8393	1.0597	14.69
	CFATransUnet [25]	0.7765	2.0769	12.85	0.9261	0.5734	14.50	0.8513	1.3252	13.68
	BDFR-B1	0.8040	1.2609	**10.32**	0.9232	0.6557	15.34	0.8636	0.9578	12.83
	BDFR-B2	**0.8066**	**1.1331**	10.70	**0.9300**	**0.4416**	**13.27**	**0.8683**	**0.7873**	**11.98**

† All results are averaged over 3 independent runs; standard deviation of mean Dice <0.5% across all benchmarks.

improves PS Dice by 3.01% over the second-best while FH gains stay modest (+0.39%), validating dual-domain boundary enhancement under low contrast. Fig. 3 confirms anatomically coherent boundaries, particularly for PS where baselines show contour fragmentation or drift.

3.3 Ablation Study

We validate each component on the IUGC 2024 dataset with the PVT-B2 backbone. As shown in Table 2, the high-frequency branch yields the largest single-component gain of +0.79% mean Dice, cutting PS HD from 13.20 to 11.76 and

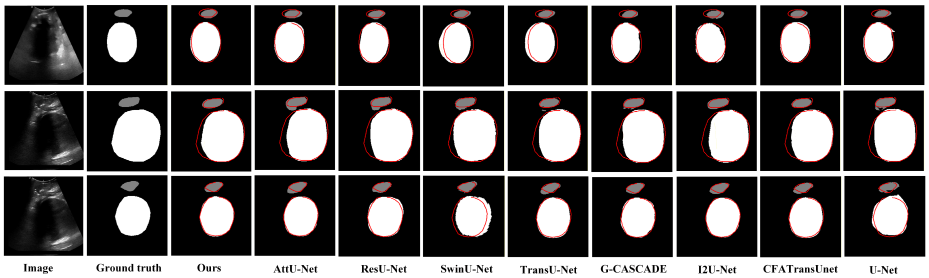

Fig. 3. Visualization of segmentation results on three randomly selected test samples. Red contours indicate ground truth annotations, while predicted boundaries of the fetal head and pubic symphysis are shown in white and gray, respectively.

Table 2. Ablation study on the IUGC 2024 dataset with PVT-B2 backbone. HF: high-frequency branch; LF: low-frequency branch; CTX: spatial context branch; ACAM: adaptive context alignment module. HF+LF+CTX collectively constitute the DDBE module. Best results in **bold**.

Components				PS			FH			Mean		
HF	LF	CTX	ACAM	Dice	ASD	HD	Dice	ASD	HD	Dice	ASD	HD
				78.10	1.48	13.20	91.37	0.87	15.90	84.74	1.17	14.55
✓				79.14	1.31	11.76	91.92	0.69	14.73	85.53	1.00	13.25
✓	✓			79.58	1.24	11.43	92.31	0.56	13.86	85.95	0.90	12.65
✓	✓	✓		80.13	1.19	10.94	92.60	0.53	13.72	86.37	0.86	12.33
			✓	78.72	1.40	12.68	91.69	0.75	15.14	85.21	1.08	13.91
✓	✓	✓	✓	**80.66**	**1.13**	**10.70**	**93.00**	**0.44**	**13.27**	**86.83**	**0.79**	**11.98**

confirming directional wavelet decomposition as the key cue for low-contrast PS boundaries. The low-frequency branch further raises mean Dice to 85.95% and lowers FH HD from 14.73 to 13.86, stabilizing large-scale contours. The spatial context branch adds +0.42% mean Dice and reduces PS HD from 11.43 to 10.94, resolving inter-regional ambiguity between the separated FH and PS. ACAM alone improves the baseline by +0.47%, and with the full DDBE yields a further +0.46% gain (86.37%→86.83%) with mean ASD from 0.86 to 0.79, confirming that cross-level alignment complements frequency-spatial enhancement.

4 Conclusion

This paper presents BDFR-Net, a boundary-aware dual-domain network that addresses anatomical boundary ambiguity in fetal ultrasound segmentation through synergistic frequency-spatial feature reconstruction. The Dual-Domain Feature Decoupling and Boundary Enhancement (DDBE) module disentangles multi-scale spectral decompositions into directional edge responses and stable structural representations, reconciling them with spatially-modeled long-range dependencies. The Adaptive Context Alignment Module (ACAM) mitigates cross-level semantic misalignment in skip connections through explicit hierarchical recalibration. Extensive experiments demonstrate state-of-the-art performance:

97.01% mean Dice on FH-PS-AoP, 82.94% on Cross-dataset, and 86.83% on IUGC 2024. The consistent improvements on the anatomically challenging pubic symphysis region validate the effectiveness of dual-domain boundary modeling under low contrast conditions. Future work will explore more lightweight model designs and extend the framework to broader clinical applications.

Acknowledgments. This work was supported by the Special Foundation for State Major Basic Research Program of China under Grant 2025YFC2708700, in part by the National Natural Science Foundation of China under Grant 62302338, in part by the Natural Science Foundation of Chongqing under Grants CSTB2025NSCQ-GPX1037 and CSTB2023NSCQ-MSX0407, in part by the Natural Science Foundation Innovation and Development Joint Fund Project of Chongqing under Grant CSTB2023NSCQ-LZX0148, in part by the Brain-Gain Plan of New Chongqing Foundation under Grant CSTB2024YCJH-KYXM0108, in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grants KJQN202500532 and KJQN202500511, in part by the Chongqing Normal University Foundation under Grant 24XLB018, and in part by the Shandong Province Natural Science Foundation under Grants ZR2022QF046 and ZR2025MS1067.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Miccai iugc 2024: Intrapartum ultrasound grand challenge (2024), https://conferences.miccai.org/2024/en/
2. Asgari, R., Waldstein, S., Schlanitz, F., Baratsits, M., Schmidt-Erfurth, U., Bogunović, H.: U-net with spatial pyramid pooling for drusen segmentation in optical coherence tomography. In: Ophthalmic Medical Image Analysis: 6th International Workshop, OMIA 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, Proceedings 6. pp. 77–85. Springer (2019)
3. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(12), 2481–2495 (2017)
4. Cai, S., Tian, Y., Lui, H., Zeng, H., Wu, Y., Chen, G.: Dense-unet: A novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. Quantitative Imaging in Medicine and Surgery **10**(6), 1275 (2020)
5. Chen, G., Bai, J., Ou, Z., Lu, Y., Wang, H.: Psfhs: intrapartum ultrasound image dataset for ai-based segmentation of pubic symphysis and fetal head. Scientific Data **11**(1), 436 (2024)
6. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
7. Chen, Z., Ou, Z., Lu, Y., Bai, J.: Direction-guided and multi-scale feature screening for fetal head–pubic symphysis segmentation and angle of progression calculation. Expert Systems with Applications **245**, 123096 (2024)
8. Dai, D., Dong, C., Yan, Q., Sun, Y., Zhang, C., Li, Z., Xu, S.: I2u-net: A dual-path u-net with rich information interaction for medical image segmentation. Medical Image Analysis **97**, 103241 (2024)

9. Dharejo, F.A., Zawish, M., Deeba, F., Zhou, Y., Dev, K., Khowaja, S.A., Qureshi, N.M.F.: Multimodal-boost: Multimodal medical image super-resolution using multi-attention network with wavelet transform. IEEE/ACM Transactions on Computational Biology and Bioinformatics **20**(4), 2420–2433 (2022)
10. Fu, H., Liang, J., Fang, Z., Han, J., Liang, F., Zhang, G.: Weconvene: Learned image compression with wavelet-domain convolution and entropy model. In: European Conference on Computer Vision. pp. 37–53. Springer (2024)
11. Gao, X., Qiu, T., Zhang, X., Bai, H., Liu, K., Huang, X., Wei, H., Zhang, G., Liu, H.: Efficient multi-scale network with learnable discrete wavelet transform for blind motion deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2733–2742 (2024)
12. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI Brainlesion Workshop. pp. 272–284. Springer (2021)
13. Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.W., Wu, J.: Unet 3+: A full-scale connected unet for medical image segmentation. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1055–1059. IEEE (2020)
14. Jordan, R.G., Cockerham, A.Z.: Common discomforts of pregnancy. Prenatal and Postnatal Care: A Person-Centered Approach **233** (2023)
15. Kalache, K.D., Dückelmann, A., Michaelis, S., Lange, J., Cichon, G., Dudenhausen, J.: Transperineal ultrasound imaging in prolonged second stage of labor with occipitoanterior presenting fetuses: how well does the 'angle of progression'predict the mode of delivery? Ultrasound in Obstetrics and Gynecology **33**(3), 326–330 (2009)
16. Li, Q., Shen, L., Guo, S., Lai, Z.: Wavelet integrated cnns for noise-robust image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7245–7254 (2020)
17. Lu, Y., Zhou, M., Zhi, D., Zhou, M., Jiang, X., Qiu, R., Ou, Z., Wang, H., Qiu, D., Zhong, M., et al.: The jnu-ifm dataset for segmenting pubic symphysis-fetal head. Data in brief **41**, 107904 (2022)
18. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
19. Rahman, M.M., Marculescu, R.: G-cascade: Efficient cascaded graph convolutional decoding for 2d medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7728–7737 (2024)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
21. Rueda, S., Fathima, S., Knight, C.L., Yaqub, M., Papageorghiou, A.T., Rahmatullah, B., Foi, A., Maggioni, M., Pepe, A., Tohka, J., et al.: Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge. IEEE Transactions on medical imaging **33**(4), 797–813 (2013)
22. Sarris, I., Ioannou, C., Chamberlain, P., Ohuma, E., Roseman, F., Hoch, L., Altman, D., Papageorghiou, A., Fetal, I., for the 21st Century (INTERGROWTH-21st), N.G.C.: Intra-and interobserver variability in fetal ultrasound measurements. Ultrasound in obstetrics & gynecology **39**(3), 266–273 (2012)

23. Tragakis, A., Kaul, C., Murray-Smith, R., Husmeier, D.: The fully convolutional transformer for medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3660–3669 (2023)
24. Valanarasu, J.M.J., Patel, V.M.: Unext: Mlp-based rapid medical image segmentation network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 23–33. Springer (2022)
25. Wang, C., Wang, L., Wang, N., Wei, X., Feng, T., Wu, M., Yao, Q., Zhang, R.: Cfatransunet: Channel-wise cross fusion attention and transformer for 2d medical image segmentation. Computers in Biology and Medicine **168**, 107803 (2024)
26. Wang, H., Cao, P., Wang, J., Zaiane, O.R.: Uctransnet: Rethinking the skip connections in u-net from a channel-wise perspective with transformer. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 2441–2449 (2022)
27. Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pvt v2: Improved baselines with pyramid vision transformer. Computational visual media **8**(3), 415–424 (2022)
28. Yao, T., Pan, Y., Li, Y., Ngo, C.W., Mei, T.: Wave-vit: Unifying wavelet and transformers for visual representation learning. In: European Conference on Computer Vision. pp. 328–345. Springer (2022)
29. Zhou, Y., Chen, H., Li, Y., Liu, Q., Xu, X., Wang, S., Yap, P.T., Shen, D.: Multitask learning for segmentation and classification of tumors in 3d automated breast ultrasound images. Medical image analysis **70**, 101918 (2021)
30. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: International Workshop on Deep Learning in Medical Image Analysis. pp. 3–11. Springer (2018)