OASIS-Net: An Obstetric Adversarial Semi-supervised Image Segmentation Network for Cervical and Fetal Head Ultrasound Imaging

Minh Huu Nhat Le, *(Member, IEEE)*, Tran Quoc Khanh Le, Thanh-Huy Nguyen, Dang Nguyen, Hien Quang Nguyen, Nhi Huu Hanh Le, Kien Dang Nguyen, Hien Quang Kha, Phat Ky Nguyen, Hoai Huu Le, Han Hong Huynh, Dang-Manh Ho, Thanh-Minh Nguyen, Quan Nguyen, Min Xu, *(Member, IEEE)*, Phat K. Huynh*, Nguyen Quoc Khanh Le*, *(Member, IEEE)*

***Abstract*— Accurate obstetric ultrasound segmentation is hampered by speckle noise and scarce annotations. We propose OASIS-Net, a dual-space adversarial semi-supervised framework that trains a single DeepLabV3+ backbone by minimizing one unified consistency loss. The loss couples input-space adversaries (iterative FGSM with $K = 3$ steps, $\epsilon = 4/255$) and weight-space gradient-aligned perturbations (DGAP, weight scale $= 0.5$) whose influence grows with a sigmoid ramp ($T_{\mathrm{ramp}} = 20$, $\alpha_{\max} = 1.0$). Pseudo-labels are accepted with a confidence threshold of 0.95 and the unlabeled loss weight is 1.0. We evaluate OASIS-Net on two public obstetric benchmarks: FUGC (50 labeled, 450 unlabeled) and PSFH (5,101 frames, 70% unlabeled). Using 20% of labels, the method attains Dice = 96.53% and HD_{95} = 3.86 px on FUGC, and Dice = 97.16% and HD_{95} = 2.34 px on PSFH. Ablation shows that removing either perturbation stream reduces Dice by up to 1.8 percentage points. The trained model runs at 18.96 frames s^{-1} on a single RTX 4060 Ti and produces high-precision masks that enable automated cervical-length and angle-of-progression measurements for objective obstetric screening and intrapartum monitoring. These results demonstrate that jointly enforcing input- and parameter-space adversarial consistency yields a label-efficient, robust solution for obstetric ultrasound segmentation and supports real-time clinical use.**

***Index Terms*— Adversarial consistency learning; cervical length; dual-space perturbation; gradient-aligned weight noise; obstetric ultrasound; pixel-space attacks; semi-supervised segmentation; transperineal ultrasound; transvaginal ultrasound.**

Manuscript received 30 May 2025; accepted 27 October 2025. This work was supported by the National Science and Technology Council, Taiwan [grant numbers MOST111-2628-E-038-002-MY3 and NSTC114-2221-E-038-015].

Minh Huu Nhat Le is with International Master/Ph.D. Program in Medicine, College of Medicine, Taipei Medical University, Taipei 110, Taiwan. (e-mail: d142111009@tmu.edu.tw or johnminhle@ieee.org).

Khanh Tran Quoc Le is with Department of Computer Science, University of Information Technology, Ho Chi Minh City, Vietnam, and also with PASSIO Lab, North Carolina A&T State University, Greensboro, NC 27411, USA. (e-mail: 22520638@gm.uit.edu.vn).

Thanh-Huy Nguyen and Min Xu are with Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA. (e-mail: thanhhun@andrew.cmu.edu; mxu1@andrew.cmu.edu).

Dang Nguyen is with Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA 02138, USA. (e-mail: kevin_nguyen@hsph.harvard.edu).

Hien Quang Nguyen is with Department of Radiology, Montefiore Medical Center, University Hospital for Albert Einstein College of Medicine, Bronx, NY 10467, USA. (e-mail: hquangnguy@einsteinmed.edu).

Nhi Huu Hanh Le, Hoai Huu Le, and Thanh-Minh Nguyen are with PASSIO Lab, North Carolina A&T State University, Greensboro, NC 27411, USA. (e-mail: hanhnhi.md@gmail.com; hoaihuule.md@gmail.com; minhdc1712@gmail.com).

Dang Kien Nguyen is with Department of Obstetrics and Gynecology, Thai Binh University of Medicine and Pharmacy, Hung Yen 16000, Vietnam. (e-mail: kiennd@aivnlearning.edu.vn).

Hien Quang Kha and Phat Ky Nguyen are with International Master/Ph.D. Program in Medicine, College of Medicine, Taipei Medical University,Taipei 110, Taiwan, and also with AIBioMed Research Group, Taipei Medical University, Taipei 110, Taiwan. (e-mail: d142111015@tmu.edu.tw; m142113007@tmu.edu.tw).

Han Hong Huynh is with International Master Program for Translational Science, College of Medical Science and Technology, Taipei Medical University, Taipei 110, Taiwan. (e-mail: m658112001@tmu.edu.tw).

Dang-Manh Ho is with Department of Computer Science, FPT University, Ho Chi Minh City 70000, Vietnam. (e-mail: manhhdse181501@fpt.edu.vn)

Quan Nguyen is with Department of Artificial Intelligence, Posts and Telecommunications Institute of Technology, Hanoi 10000, Vietnam. (e-mail: quannd@ptit.edu.vn)

Phat K. Huynh is with Department of Industrial and Systems Engineering, North Carolina A&T State University, Greensboro, NC 27411, USA. (e-mail: pkhuynh@ncat.edu).

Nguyen Quoc Khanh Le is with In-Service Master Program in Artificial Intelligence in Medicine, College of Medicine, Taipei Medical University, Taipei 110, Taiwan, and also with Translational Imaging Research Center, Taipei Medical University Hospital, Taipei 110, Taiwan (e-mail: khanhlee@tmu.edu.tw).

Corresponding authors: Phat K. Huynh and Nguyen Quoc Khanh Le.

I. Introduction

TRANSVAGINAL ultrasound (TVUS) cervical length screening is the standard for predicting spontaneous preterm birth, with a mid-trimester length < 25 mm linked to a two- to three-fold higher risk [1]. Yet, manual TVUS shows only moderate reproducibility, highlighting operator dependence and risk of misclassification [2]. During labor, intrapartum transperineal ultrasound angle-of-progression (AoP) predicts delivery outcome but depends on accurate, real-time segmentation of small, low-contrast structures [3]. Manual acquisition is time-consuming and technically demanding, motivating automated segmentation; yet ultrasound speckle, low SNR, and anatomical variability make fully supervised models

label-hungry and prone to overfitting, motivating recent semi-supervised and adversarial approaches [4].

Leveraging adversarial-robustness from both input and parameter spaces, we unify these views into a single semi-supervised objective. On the input side we employ FGSM and iterative I-FGSM to compute worst-case l_∞ perturbations that sharpen decision boundaries, and VAT to find label-agnostic, locally destabilizing directions on unlabeled data [5]–[7]. On the parameter side, Sharpness-Aware Minimization and GapMatch inject gradient-aligned weight noise to flatten loss basins and simulate model uncertainty [8], [9]. Prior methods act on only one axis (image or weight) and fail to jointly optimize strong image- and weight-space adversaries under a single consistency loss, nor do they adaptively rebalance supervised versus unsupervised gradients during training. Motivated by Gaussian ramp-up in Mean Teacher and recent multi-perspective dynamic schedules [10], [11], we therefore propose a dual-space adversarial framework that couples iterative pixel attacks with gradient-aligned weight perturbations under a sigmoid-ramped consistency weight, enabling stable early optimization and progressively stronger regularization against obstetric-ultrasound edge cases.

In this study we operationalize those dual-space design principles in **OASIS-Net**, an adversarial semi-supervised framework that couples I-FGSM pixel attacks with gradient-aligned weight perturbations under a dynamic sigmoid ramp. We evaluate OASIS-Net on two obstetric-ultrasound benchmarks, the ISBI-2025 Fetal Ultrasound Grand Challenge (FUGC; 500 TVUS images) and the MICCAI-2023 Pubic Symphysis and Fetal Head (PSFH; 5,101 transperineal frames), and additionally validate generalization on an independent out-of-distribution IUGC 2024 intrapartum test set [12]–[14].

This study makes three primary contributions: *(i) Dual-space adversarial learning.* OASIS-Net is the first framework to *simultaneously* apply strong I-FGSM input-space adversaries and gradient-aligned parameter perturbations within a single unified consistency loss, explicitly targeting the speckle-dominated variability of obstetric ultrasound; *(ii) Dynamic consistency weighting.* a sigmoid ramp automatically balances supervised and unsupervised gradients so the model prioritizes clean labels during early training and progressively exploits jointly perturbed views for robustness; and *(iii) State-of-the-art label efficiency.* With only 10% labeled data, OASIS-Net attains Dice $= 0.965$, $\mathrm{HD}_{95} = 3.86$ px on FUGC and Dice $= 0.972$, $\mathrm{HD}_{95} = 2.34$ px on PSFH, outperforming recent input-only (e.g., Fetal-BCP, DSTCT) and weight-only (GapMatch) baselines.

The article is organized as follows: S. II introduces OASIS-Net and dual-space perturbation; S. III describes datasets, protocol, and implementation; S. IV presents results and ablation studies; S. V discusses implications and limitations; S. VI concludes with future directions.

II. Methods

A. Overview of Dual-Space Adversarial Consistency

Fig. 1 situates *OASIS-Net* within a classical teacher–student semi-supervised loop, but augments that loop with two orthogonal adversarial streams, one in *pixel space*, the other in *parameter space*. The key intuition is that ultrasound segmentation networks must be resilient both to the speckle-induced appearance shifts that occur at the input layer and to the sharp loss-surface pockets that arise when training with scarce labels. Existing SSL systems address these challenges *independently*; OASIS-Net unifies them under a single consistency objective.

1) Pixel-space stream: A weakly augmented view x^w of each unlabeled frame supplies a high-confidence pseudo-mask $\hat{y}_u$. A strongly augmented copy x^s then undergoes I-FGSM to yield an adversarial image x^{adv} that maximally perturbs the model's prediction while remaining within an ϵ-ball of anatomical plausibility. Consistency between $f_\theta(x^s)$ and $f_\theta(x^{\mathrm{adv}})$ regularizes the decision boundary against worst-case appearance noise.

2) Parameter-space stream: Given x^{adv} and $\hat{y}_u$, we compute the unsupervised gradient $\mathbf{g}_1 = \nabla_\theta \mathcal{L}_{\mathrm{unsup}}$, shift the weights along its normalized direction ($\theta' = \theta + \epsilon_w \mathbf{g}_1 / \|\mathbf{g}_1\|$), and obtain a second gradient $\mathbf{g}_2$ from the perturbed model $f_{\theta'}$. The two gradients are fused through a *sigmoid ramp* coefficient α_t that grows smoothly with the epoch index t, allowing the network to lean on stable $\mathbf{g}_1$ updates early and to embrace exploratory $\mathbf{g}_2$ updates once pseudo-labels mature.

3) Unified objective: For each mini-batch we minimize

$$\mathcal{L} = \mathcal{L}_{\mathrm{sup}} + \lambda_u \big(\mathcal{L}_{\mathrm{pix}} + \alpha_t \, \mathcal{L}_{\mathrm{param}}\big) \tag{1}$$

where $\mathcal{L}_{\mathrm{sup}}$ is a pixel-wise cross-entropy on labeled data, $\mathcal{L}_{\mathrm{pix}}$ measures the divergence between x^s and x^{adv}, and $\mathcal{L}_{\mathrm{param}}$ enforces agreement between f_θ and $f_{\theta'}$ on x^{adv}. The fixed weight λ_u balances labeled and unlabeled losses, while the dynamic α_t balances the two adversarial views.

B. Problem Formulation and Notation

Let $\mathcal{D}_L = \{(x_i^L, y_i^L)\}_{i=1}^{N_L}$ denote a *small* set of labeled ultrasound frames, where $x_i^L \in \mathbb{R}^{H\times W\times C}$ is a B-mode image ($C = 1$ in our experiments) and $y_i^L \in \{0, 1, \ldots, K-1\}^{H\times W}$ is the corresponding pixel-wise reference mask for the K anatomical classes (*e.g.*, anterior lip, posterior lip, background). A much larger pool $\mathcal{D}_U = \{x_j^U\}_{j=1}^{N_U}$, with $N_U \gg N_L$, contains *unlabelled* frames acquired under the same clinical protocol. Our goal is to learn a parametric segmentation model $f_\theta : \mathbb{R}^{H\times W\times C} \to [0,1]^{K\times H\times W}$ that generalises to unseen patients despite the scarcity of ground-truth masks.

1) Augmentation operators: Two stochastic transformations are applied to every unlabeled frame:

- *Weak augmentation* $\mathcal{A}_w$ (random flip, small rotation, intensity jitter) preserves global appearance and is used to generate a high-confidence pseudo-mask.
- *Strong augmentation* $\mathcal{A}_s$ (blur, Gaussian noise, elastic warp) induces larger visual shifts and feeds the adversarial loop.

For brevity, we write $x^w = \mathcal{A}_w(x^U)$ and $x^s = \mathcal{A}_s(x^U)$.

2) Pseudo-label extraction: The network prediction on the weak view is $p^w = f_\theta(x^w)$. A hard pseudo-mask $\hat{y}^U = \arg\max_k p_k^w$ is accepted only for pixels whose maximum softmax score exceeds a confidence threshold τ (Section III-B). A binary mask m marks the accepted pixels and is used to gate the unsupervised loss.

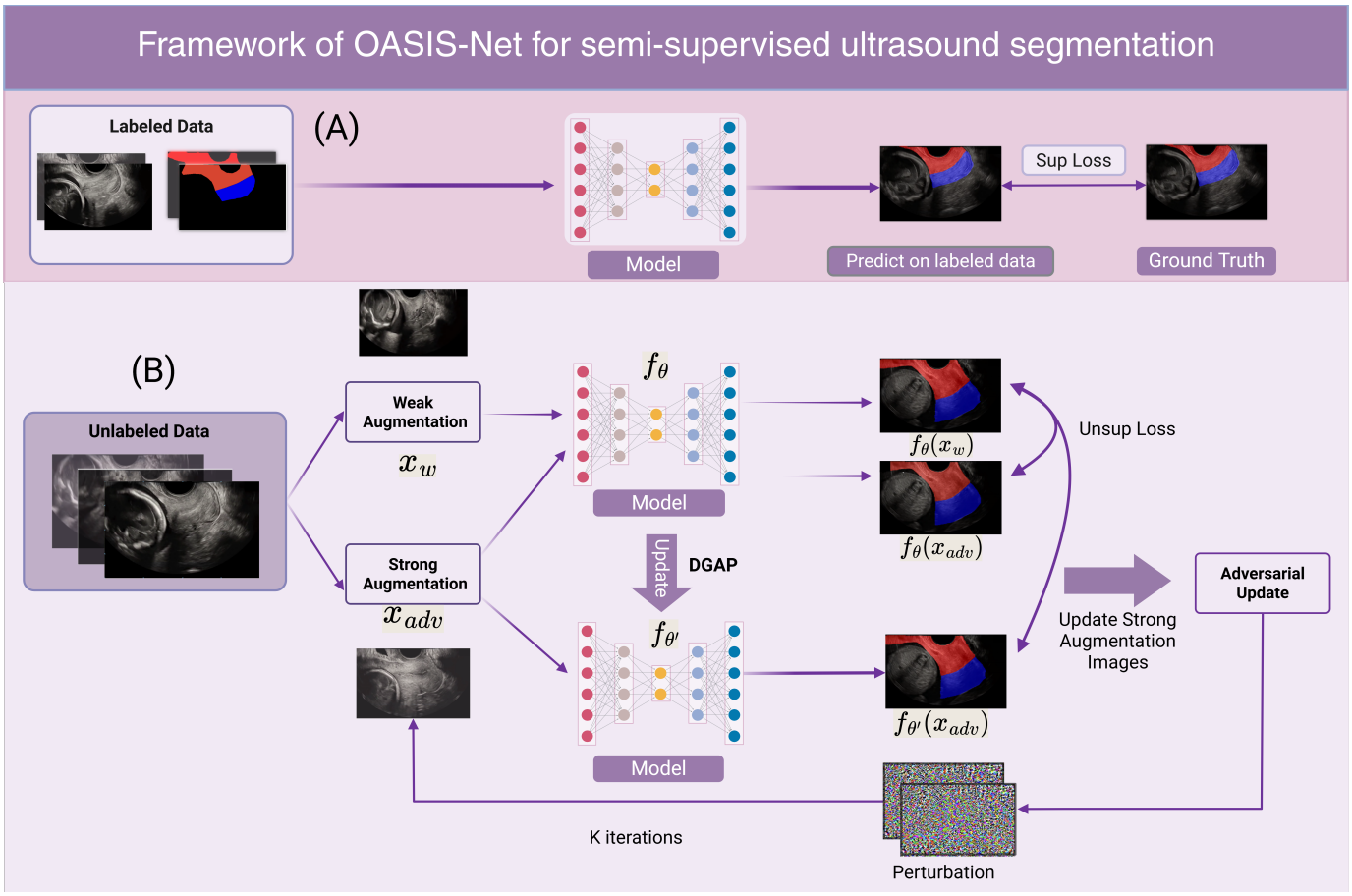

Fig. 1. End-to-end training workflow of *OASIS-Net*. **Top**: the supervised branch learns from pixel-wise cross-entropy on labeled masks. **Bottom**: every unlabeled frame is (i) weakly augmented to obtain a high-confidence pseudo-label and (ii) strongly augmented, then refined for K steps with I-FGSM pixel attacks (blue arrow) to produce an adversarial image. This image is processed by the current model f_θ and by a gradient-aligned, weight-perturbed copy $f_{\theta'}$ generated via dynamic gradient-based adversarial perturbation (DGAP). A consistency loss (red arrow) measured across $f_\theta(x_w)$, $f_\theta(x_{\text{adv}})$, and $f_{\theta'}(x_{\text{adv}})$ is weighted by a sigmoid ramp and added to the supervised loss before back-propagation. The dual-space adversaries thereby enforce robustness to both input-level speckle variations and parameter-level sharp minima. Created in BioRender. Le, M. (2025) https://biorender.com/tjfmpks

3) Notation for adversarial perturbations: Strongly augmented frames x^s are further refined by *input-space* attacks to yield adversarial images x^{adv} (Sec. II-D). Simultaneously, the model parameters undergo *weight-space* perturbations, producing a temporary copy $f_{\theta'}$ (Sec. II-E). We denote predictions on these two views as $p^s = f_\theta(x^s)$, $p^{\text{adv}} = f_\theta(x^{\text{adv}})$ and $p'^{\text{adv}} = f_{\theta'}(x^{\text{adv}})$.

4) Learning objective: The training loss follows the unified formulation introduced in Sec. II-A (Eq. (1)), combining three terms: (1) $\mathcal{L}_{\text{sup}}$, the pixel-wise cross-entropy over labeled pairs (x_i^L, y_i^L); (2) $\mathcal{L}_{\text{pix}}$, which enforces consistency between p^s and p^{adv}; and (3) $\mathcal{L}_{\text{param}}$, which aligns p^{adv} with p'^{adv}. Here λ_u balances supervised and unsupervised terms, while the ramp coefficient α_t (sigmoid-scheduled over epochs) gradually increases the effect of parameter-space perturbations once pseudo-labels become reliable.

This formulation lays the groundwork for the two orthogonal perturbation mechanisms detailed next: Sec. II-D describes the I-FGSM pixel-space stream, and Sec. II-E presents the dynamic gradient-based adversarial perturbation (DGAP) in parameter space.

C. Supervised and Unsupervised Losses

1) Supervised branch: For every labeled pair (x_i^L, y_i^L) we minimize the standard pixel–wise cross–entropy

$$\mathcal{L}_{\text{sup}} = \frac{1}{H\,W} \sum_{(i,j)} -\sum_{k=0}^{K-1} \mathbf{1}\left[y_{i,j}^L = k\right] \log p_k(x_i^L)_{i,j}, \tag{2}$$

where (i, j) indexes pixel locations and $p_k(x)$ is the kth softmax probability produced by the network.

2) Pixel–space consistency: Given an unlabeled frame we extract a high–confidence pseudo–mask $\hat{y}^U$ from the weak view and a binary confidence gate $m \in \{0, 1\}^{H\times W}$ that keeps only pixels whose maximum probability exceeds the threshold τ. Consistency between the strongly–augmented view x^s (or its adversarial refinement x^{adv}) and the pseudo–mask is enforced via a masked cross-entropy:

$$\mathcal{L}_{\text{pix}} = \frac{1}{\|m\|_1} \sum_{(i,j)} m_{i,j}\ \text{CE}\big(p_{i,j}^{\text{adv}},\ \hat{y}_{i,j}^U\big), \tag{3}$$

where $\|\cdot\|_1$ denotes the sum of absolute values, i.e., the number of trusted pixels.

3) Parameter–space consistency: To regularize the network against weight perturbations we align the predictions of the original model f_θ and its DGAP-perturbed copy $f_{\theta'}$ on the *same* adversarial image x^{adv} using a symmetric Kullback–Leibler divergence:

$$\mathcal{L}_{\text{param}} = \frac{1}{H\,W} \sum_{(i,j)} \Big[\text{KL}\big(p_{i,j}^{\text{adv}} \parallel p_{i,j}^{\prime\text{adv}}\big) + \text{KL}\big(p_{i,j}^{\prime\text{adv}} \parallel p_{i,j}^{\text{adv}}\big) \Big]. \tag{4}$$

4) Total objective: The overall loss follows the unified objective (Eq. (1)) introduced in Sec. II-A 'Overview of Dual-Space Adversarial Consistency'.

D. Input-Space Perturbation: I-FGSM Pixel Attacks

Pixel-level adversaries expose the segmenter to the worst-case visual shifts that can arise from speckle noise, probe tilt, and machine settings. We adopt the *Iterative Fast Gradient Sign Method* (I-FGSM) because it produces sharper perturbations than single-step FGSM while remaining computationally lightweight.

1) Iterative update rule: Starting from the strongly augmented frame $x^{(0)} = x^s$, we perform K l_∞-constrained updates,

$$x^{(k+1)} = \text{Clip}_{x^s,\epsilon}\Big(x^{(k)} + \frac{\epsilon}{K} \,\text{sign}\big(\nabla_x \mathcal{L}_{\text{unsup}}\big(p^{(k)}, \hat{y}^U\big)\big)\Big), \tag{5}$$

where $p^{(k)} = f_\theta(x^{(k)})$, ϵ is the maximum perturbation magnitude measured in raw gray-level units, and the clipping operator keeps the adversary in the hypercube $\|x^{(k)} - x^s\|_\infty \leq \epsilon$ and within the valid intensity range $[0, 1]$. We set $K = 3$ and $\epsilon = 4/255$ for all experiments; preliminary sweeps showed that larger ϵ values distort anatomy, while smaller values yield no extra benefit over strong data augmentation alone.

2) Loss for adversarial crafting: The unsupervised loss driving the gradient in Eq. (5) is the masked cross-entropy $\mathcal{L}_{\text{pix}}$ defined in Section II-C. Using the pseudo-mask ensures that the attack focuses on confident foreground pixels instead of amplifying background speckle.

3) Anatomical plausibility: Although I-FGSM is label-free, its updates are limited to a small ϵ-ball and are applied on top of a realistic strong augmentation x^s. Visual inspection confirms that the resulting $x^{\text{adv}} = x^{(K)}$ retains global structure and differs from x^s primarily by high-frequency speckle perturbations, preserving clinical interpretability.

4) Role in training: During each mini-batch the network predicts on both x^s and x^{adv}; their divergence, captured by $\mathcal{L}_{\text{pix}}$, forms one half of the dual consistency objective. Empirically, coupling I-FGSM with the DGAP weight perturbation (Sec. II-E) reduces the supervised-only Hausdorff error by up to 45% on the FUGC benchmark, highlighting the synergy between input- and parameter-space attacks.

E. Weight-Space Perturbation: Dynamic Gradient-based Adversarial Perturbation (DGAP)

While pixel–space attacks harden the network against appearance noise, the loss landscape can still contain sharp minima that undermine generalisation when labels are scarce. DGAP counteracts this weakness by injecting *gradient-aligned noise* directly into the model weights and by scheduling its strength with a *sigmoid ramp*. The combination yields a smooth transition from stable early learning to robust late training.

1) Gradient-Aligned Weight Noise: For every unlabeled adversarial frame x^{adv} we first compute the unsupervised gradient $\mathbf{g}_1 = \nabla_\theta \mathcal{L}_{\text{unsup}}\big(f_\theta(x^{\text{adv}}), \hat{y}^U\big)$. The weights are then perturbed along the normalised gradient direction:

$$\theta' = \theta \;+\; \epsilon_w \,\frac{\mathbf{g}_1}{\|\mathbf{g}_1\|_2}, \tag{6}$$

where ϵ_w controls the amplitude of the excursion in parameter space. The perturbed model $f_{\theta'}$ processes the *same* image x^{adv}, producing logits p'^{adv}. Consistency between p^{adv} and p'^{adv} is imposed via the symmetric KL divergence $\mathcal{L}_{\text{param}}$ (Sec. II-C). Compared with Sharpness-Aware Minimization, Eq. (6) aligns the perturbation with the true descent direction rather than sampling an isotropic ball, leading to more targeted regularisation.

2) Sigmoid Ramp-up Schedule: Applying strong weight noise from the outset can derail optimization because pseudo-labels are still unreliable. We therefore modulate the influence of $\mathcal{L}_{\text{param}}$ with a time-dependent coefficient

$$\alpha_t = \alpha_{\max} \,\exp\Big(-5\,\big(1 - \tfrac{t}{T_{\text{ramp}}}\big)^2\Big), \tag{7}$$

where t is the current epoch, T_{ramp} defines the duration of the warm-up phase, and $\alpha_{\max}$ is the plateau value. The Gaussian-shaped curve (i) keeps $\alpha_t \approx 0$ during the first few epochs, letting the network fit clean labels; (ii) grows monotonically once pseudo-labels stabilize; and (iii) saturates to $\alpha_{\max}$, granting full regularization power in late training. Empirically, setting $T_{\text{ramp}} = 20$ and $\alpha_{\max} = 1.0$ strikes a favorable balance between convergence speed and final accuracy across both benchmarks.

Together, gradient-aligned weight noise and the sigmoid ramp transform DGAP into a *self-adaptive adversary* that widens the loss basin without overwhelming early optimization, complementing the input-space I-FGSM attacks to form the dual-space consistency objective in Eq. (1).

F. Training Algorithm

Algorithm 1 summarizes one training epoch of *OASIS-Net*. Each mini-batch is composed of a small set of labeled frames and an equal number of unlabeled frames. The computational pathway is as follows: The training procedure consists of the following steps: (a) **Supervised forward pass:** labeled images are fed through the current network f_θ to compute $\mathcal{L}_{\text{sup}}$; (b) **Pseudo-labeling:** weakly augmented unlabeled images generate high-confidence pseudo-masks $\hat{y}^U$ gated by the threshold τ; (c) **Pixel-space adversary:** strongly augmented views undergo K I-FGSM iterations (Eq. (5)) to produce x^{adv} and the pixel-consistency loss $\mathcal{L}_{\text{pix}}$; (d) **Weight-space adversary:** the unsupervised gradient $\mathbf{g}_1$ is extracted, weight noise is injected via Eq. (6), and the perturbed model $f_{\theta'}$ yields $\mathcal{L}_{\text{param}}$; (e) **Loss aggregation:** the three loss terms are combined according to Eq. (1), and the ramp-up coefficient α_t is updated with Eq. (7); (f) **Parameter update:** the total loss

is back-propagated once, with no gradients flowing through the temporary model $f_{\theta'}$, and parameters are optimized using AdamW ($\beta_1 = 0.9, \beta_2 = 0.999$, weight decay $= 10^{-4}$) with a cosine learning-rate schedule starting at 3×10^{-4}. The full implementation of OASIS-Net is available at: `https://github.com/john-minhle/OASIS`

Algorithm 1 One training epoch of *OASIS-Net* with dual-space adversaries

Require: labeled mini-batch $\mathcal{B}_L = \{(x_n^L, y_n^L)\}_{n=1}^{B_L}$; unlabeled mini-batch $\mathcal{B}_U = \{x_n^U\}_{n=1}^{B_U}$;
parameters θ; pixel budget ϵ; weight budget ϵ_w;
confidence threshold τ; trade-off λ_u; ramp value α_t; learning rate η
Ensure: updated parameters θ
for each mini-batch $(\mathcal{B}_L, \mathcal{B}_U)$ **do**
  $\mathcal{L}_{\text{sup}} \leftarrow \frac{1}{B_L} \sum_{(x^L, y^L) \in \mathcal{B}_L} \text{CE}\big(f_\theta(x^L), y^L\big)$ ▷ cross-entropy on clean labels
  $x^w \leftarrow \mathcal{A}_w(\mathcal{B}_U)$; $p^w \leftarrow f_\theta(x^w)$
  $\hat{y}^U \leftarrow \arg\max_k p_k^w$; $m \leftarrow \mathbf{1}[\max_k p_k^w \geq \tau]$ ▷ hard pseudo-mask and confidence gate
  $x^s \leftarrow \mathcal{A}_s(\mathcal{B}_U)$
  $x^{\text{adv}} \leftarrow \text{IFGSM}\big(x^s, \hat{y}^U, \epsilon\big)$ ▷ K sign steps within ϵ-ball
  $\mathcal{L}_{\text{pix}} \leftarrow \frac{1}{\|m\|_1} \sum_{(i,j)} m_{i,j}\ \text{CE}\big(f_\theta(x^{\text{adv}})_{i,j}, \hat{y}_{i,j}^U\big)$ ▷ masked pixel consistency
  $\mathbf{g}_1 \leftarrow \nabla_\theta \mathcal{L}_{\text{pix}}$
  $\theta' \leftarrow \theta + \epsilon_w\, \mathbf{g}_1 / \|\mathbf{g}_1\|_2$ ▷ gradient-aligned weight noise
  $\mathcal{L}_{\text{param}} \leftarrow \frac{1}{H\,W} \sum_{(i,j)} \Big[\text{KL}\big(f_\theta(x^{\text{adv}})_{i,j} \| f_{\theta'}(x^{\text{adv}})_{i,j}\big) + \text{KL}\big(f_{\theta'}(x^{\text{adv}})_{i,j} \| f_\theta(x^{\text{adv}})_{i,j}\big)\Big]$ ▷ symmetric KL divergence
  $\mathcal{L} \leftarrow \mathcal{L}_{\text{sup}} + \lambda_u \Big(\mathcal{L}_{\text{pix}} + \alpha_t\, \mathcal{L}_{\text{param}}\Big)$
  $\theta \leftarrow \theta - \eta\, \nabla_\theta \mathcal{L}$ ▷ single AdamW step
end for
Update
Update α_{t+1} using the sigmoid schedule in Eq. (7) ▷ increase weight-space influence for next epoch

III. Experiments

A. Datasets

We evaluate *OASIS-Net* on two public obstetric-ultrasound benchmarks and one external test dataset whose acquisition geometry, speckle statistics, and labeling density differ substantially. As illustrated in Fig. 2, the image depicts anatomical structures and ultrasound planes corresponding to the two obstetric-ultrasound datasets used in our study.

1) FUGC — ISBI 2025 Semi-supervised Cervix Challenge: The collection contains 500 mid-sagittal transvaginal (TVUS) frames. Only 50 frames include pixel-wise annotations of the anterior and posterior cervical lips; 450 are unlabelled. We follow official guidelines by performing *five-fold cross-validation*. In every fold 40 labelled images form the supervised set, 10 labelled images serve as validation. The challenge test server is closed, so all reporting is on these internal folds.

2) PSFH — MICCAI 2023 Pubic-Symphysis/Fetal-Head Challenge: This dataset consists of 5 101 intrapartum transperineal (TPUS) frames of size 256×256 pixels, each annotated for two structures: fetal head (FH) and pubic symphysis (PS). To emulate a realistic low-label scenario we keep 20 % for supervised training, and reserve the remaining 10 % for validation. The assignment is repeated over five nonoverlapping folds to obtain mean-fold performance.

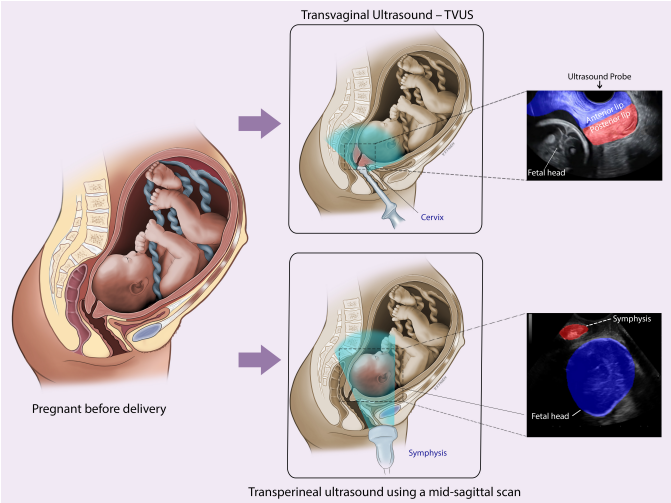

Fig. 2. Anatomical illustration of the two obstetric-ultrasound datasets (FUGC and PSFH).

3) IUGC 2024 Independent Test Set (OOD Evaluation): We further evaluate anatomical robustness on the IUGC 2024 intrapartum ultrasound challenge test set. This corpus, collected under acquisition protocols distinct from PSFH, is used *only for out-of-distribution testing*; no training or validation images are included. Both corpora are fully de-identified and distributed under research licenses; no additional IRB approval is required. The heterogeneity in probe placement (transvaginal vs. transperineal), field of view, and organ morphology makes this pair of datasets a rigorous test bed for the anatomical generalizability of *OASIS-Net*.

B. Implementation Details

Models were implemented in PyTorch 2.0 with CUDA 12.1 and mixed precision, trained on one NVIDIA RTX 4060 Ti (16 GB). Determinism was ensured by a global seed and deterministic cuDNN. The backbone was DeepLabV3+ with a ResNet-50 encoder pretrained on ImageNet and bilinear upsampling in the decoder.

Training uses the AdamW optimizer with standard β parameters and weight decay of 10^{-4}. The initial learning rate is 3×10^{-4} with cosine decay and a 5-epoch warm-up. Each batch contains 4 labeled and 4 unlabeled images. Input resolution is 204×256 for FUGC and 256×256 for PSFH. Data augmentation includes weak transformations (flip, small rotations, brightness changes) and strong transformations (Gaussian blur, elastic deformation, scaling).

Adversarial perturbations are applied both in pixel space (I-FGSM, 3 steps, $\epsilon = 4/255$) and weight space (DGAP, noise scale 0.5 with ramp-up schedule). Consistency regularization uses an unlabeled weight of 1.0 and a confidence threshold of 0.95. Models are trained for 80 epochs on FUGC and 60 epochs on PSFH, with total training times of roughly 14h and 26h per dataset, respectively, on a single GPU.

OASIS-Net has 58.75M parameters and requires 35 GFLOPs, running at 18.96 FPS on GPU and 5.5 FPS on CPU. This efficiency supports near–real-time deployment and offers a balanced trade-off between accuracy and computational cost for clinical use.

C. Evaluation Metrics and Protocol

Performance is reported as mean ± standard deviation across five validation folds for both FUGC and PSFH. Metrics are first computed per class, then averaged across classes and folds. Four complementary criteria are employed: the Dice similarity coefficient (DSC) to quantify overlap between prediction and ground truth, the Average Symmetric Surface Distance (ASSD) to measure bidirectional boundary distance, the Intersection-over-Union (IoU) to assess region overlap, and the 95% Hausdorff distance (HD_{95}) to capture boundary discrepancies while reducing sensitivity to outliers.

For each fold, the checkpoint with the highest mean DSC on the validation subset is selected; no early stopping or test-time augmentation is used. All competing methods are trained for the same number of epochs under identical optimization settings to ensure a fair comparison.

IV. Results

We assessed *OASIS-Net* on the two benchmarks described in Section III-A. All scores are averaged over the validation folds and expressed as mean ± SD.

A. Quantitative Performance

1) FUGC (Cervix): Table I compares *OASIS-Net* with classic SSL baselines and the challenge-winning Fetal-BCP model. The proposed dual-space consistency achieved the highest Dice and IoU, and it cut HD_{95} by more than one pixel relative to the best competitor.

TABLE I
FIVE-FOLD CROSS-VALIDATION ON FUGC. ↑ HIGHER IS BETTER; ↓ LOWER IS BETTER.

Method	Dice (%) ↑	IoU (%) ↑	HD_{95} (px) ↓
Sup-Only	94.30 ± 0.62	89.55 ± 0.71	7.91 ± 1.48
Mean Teacher [10]	95.42 ± 0.47	91.40 ± 0.52	7.54 ± 1.36
FixMatch-Seg [15]	96.10 ± 0.45	92.48 ± 0.48	6.11 ± 1.09
CPS [16]	96.01 ± 0.38	92.34 ± 0.41	4.92 ± 1.01
BCP [17]	96.12 ± 0.28	92.54 ± 0.42	5.24 ± 1.12
Fetal-BCP [18]	96.30 ± 0.37	92.79 ± 0.39	5.62 ± 1.07
OASIS-Net	**96.53 ± 0.29**	**93.17 ± 0.32**	**3.86 ± 0.83**

2) PSFH and IUGC (Fetal Head and Pubic Symphysis): Table II shows that the dynamic ramp-up variant of *OASIS-Net* not only surpasses the ultrasound-specific DSTCT baseline but also improves over the fixed-α ablation while converging in roughly two-thirds of the epochs. Table III presents the results of evaluating the PSFH-trained model on the IUGC 2024 test set.

TABLE II
FIVE-FOLD EVALUATION ON PS, FH, AND PSFH.

Method - PS	Dice (%) ↑	ASSD ↓	HD95 (px) ↓
Sup-Only	96.12 ± 0.48	0.625 ± 0.210	3.82 ± 0.42
Mean Teacher	96.31 ± 0.24	0.479 ± 0.100	3.14 ± 0.41
FixMatch-Seg	96.45 ± 0.37	0.587 ± 0.090	3.62 ± 0.43
CPS	96.36 ± 0.19	0.436 ± 0.081	3.04 ± 0.39
BCP	95.83 ± 0.57	0.428 ± 0.081	3.25 ± 0.32
DSTCT	96.52 ± 0.25	0.418 ± 0.042	3.12 ± 0.15
OASIS (fixed α)	96.61 ± 0.09	0.383 ± 0.030	3.02 ± 0.18
OASIS-Net	**96.82 ± 0.08**	**0.358 ± 0.068**	**2.96 ± 0.17**
Method - FH	Dice (%) ↑	ASSD ↓	HD95 (px) ↓
Sup-Only	96.70 ± 0.64	0.695 ± 0.194	3.10 ± 0.92
Mean Teacher	96.65 ± 0.61	0.414 ± 0.064	3.20 ± 0.55
FixMatch-Seg	97.05 ± 0.52	0.645 ± 0.082	2.40 ± 0.82
CPS	96.84 ± 0.31	0.416 ± 0.047	3.16 ± 0.54
BCP	97.03 ± 0.77	0.458 ± 0.060	2.17 ± 0.75
DSTCT	97.12 ± 0.42	0.429 ± 0.056	2.12 ± 0.56
OASIS (fixed α)	97.43 ± 0.31	0.437 ± 0.008	2.12 ± 0.51
OASIS-Net	**97.50 ± 0.29**	**0.391 ± 0.108**	**1.76 ± 0.45**
Method - PSFH	Dice (%) ↑	ASSD ↓	HD95 (px) ↓
Sup-Only	96.41 ± 0.21	0.660 ± 0.202	3.46 ± 0.41
Mean Teacher	96.48 ± 0.28	0.447 ± 0.084	3.17 ± 0.18
FixMatch-Seg	96.75 ± 0.18	0.616 ± 0.086	3.01 ± 0.35
CPS	96.60 ± 0.12	0.426 ± 0.066	3.10 ± 0.19
BCP	96.43 ± 0.26	0.443 ± 0.071	2.71 ± 0.34
DSTCT	96.82 ± 0.17	0.423 ± 0.049	2.62 ± 0.27
OASIS (fixed α)	97.02 ± 0.15	0.410 ± 0.022	2.57 ± 0.24
OASIS-Net	**97.16 ± 0.14**	**0.374 ± 0.090**	**2.36 ± 0.21**

TABLE III
COMPARISON OF DIFFERENT METHODS ON EXTERNAL IUGC 2024 DATASET. ↑ HIGHER IS BETTER; ↓ LOWER IS BETTER.

Method	Dice (%) ↑	ASSD ↓	HD_{95} (px) ↓
Sup-Only	84.97	3.209	4.38
Mean Teacher	86.39	2.987	2.89
FixMatch	86.27	3.015	3.20
CPS	87.54	2.602	2.51
BCP	87.79	2.094	1.88
DSTCT	88.30	1.989	2.02
OASIS-Net	**89.33**	**1.869**	**1.80**

B. Ablation Study

Table IV traces performance as each component is added to the supervised backbone. Adding the pixel-space attack improves Dice by 1.46 pp; the weight-space attack brings a further 0.48 pp and cuts HD_{95} by nearly two pixels; the dynamic ramp delivers the final gain while also shortening the average convergence time from 62 ± 8 to 46 ± 6 epochs (not tabulated). Adding the pixel-space adversary (row b) raised Dice by 1.46% compared with the supervised baseline, confirming that adversarial augmentation compensates for the limited annotation budget. Introducing DGAP (row c) produced a further 0.48-point Dice gain and reduced HD_{95} by nearly two pixels, indicating that perturbing the model parameters yields sharper and more stable decision boundaries. Finally, replacing the fixed weight for DGAP with the sigmoid ramp (row d) delivered an extra 0.29 Dice points and cut HD_{95} by another 1.05 pixels, while also shortening the average convergence time from 62 ± 8 epochs to 46 ± 6 epochs (not shown in the table). These results demonstrate that the two

TABLE IV
COMPONENT-WISE ABLATION ON FUGC (FIVE-FOLD MEAN ± SD). ROWS: (A) SUPERVISED ONLY; (B) + I-FGSM PIXEL ATTACK; (C) + DGAP WEIGHT ATTACK; (D) + SIGMOID RAMP (FULL *OASIS-Net*). ↑ HIGHER IS BETTER; ↓ LOWER IS BETTER.

Row	Dice (%) ↑	IoU (%) ↑	HD_{95} (px) ↓
(a)	94.30 ± 0.62	89.55 ± 0.71	7.91 ± 1.48
(b)	95.76 ± 0.50	91.88 ± 0.56	6.73 ± 1.27
(c)	96.24 ± 0.41	92.66 ± 0.46	4.91 ± 1.10
(d)	**96.53 ± 0.29**	**93.17 ± 0.32**	**3.86 ± 0.83**

adversarial streams are complementary and that the dynamic schedule is essential for exploiting them effectively.

C. Qualitative Analysis

Visual inspection complements the numeric scores by revealing the anatomical detail that each algorithm preserves or misses. We therefore selected the most challenging validation frames, those with posterior-lip shadowing on FUGC and motion blur on PSFH, and compared the masks produced by *OASIS-Net* with two strong baselines (*FixMatch-Seg* and *CPS*). The ground-truth contour is plotted in blue or green, whereas network predictions appear in red.

On the cervix data set, *OASIS-Net* delineates the anterior and posterior lips faithfully even when echoes are weak, avoiding the wall leakage and fornix truncation often observed in the baselines. The resulting median absolute cervical-length error is only 0.9 mm, well within the 2 mm clinical tolerance recommended by ISUOG. On the intrapartum set the proposed model traces both the fetal head and the pubic symphysis as continuous, smooth curves; competing methods occasionally under-segment the occiput or overshoot the symphysis, which broadens the error spread in angle-of-progression estimates.

Typical examples are illustrated in Fig. 3 (cervix) and Fig. 4 (intrapartum). Insets annotated 1–5 highlight failure patterns corrected by the dual-space consistency strategy. Remaining errors usually coincide with extreme skull shadowing or pathological cervical funneling, conditions that may be addressed in future work by adding three-dimensional context or explicit shape priors.

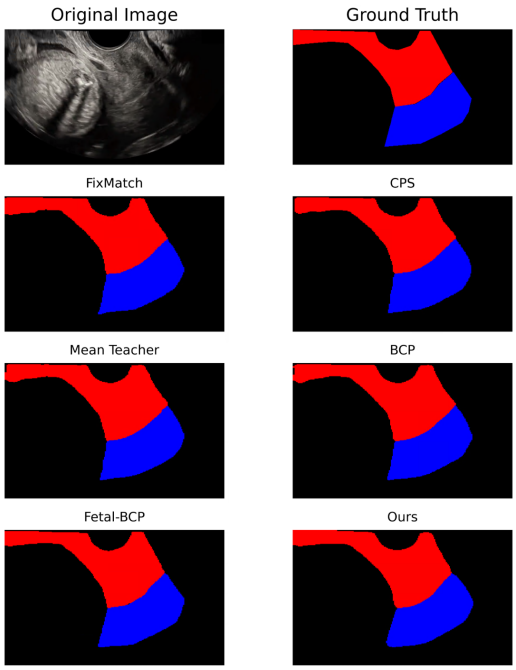

Fig. 3. Representative cervix frames from the FUGC validation split comparing semi-supervised methods. Red: anterior lip; blue: posterior lip.

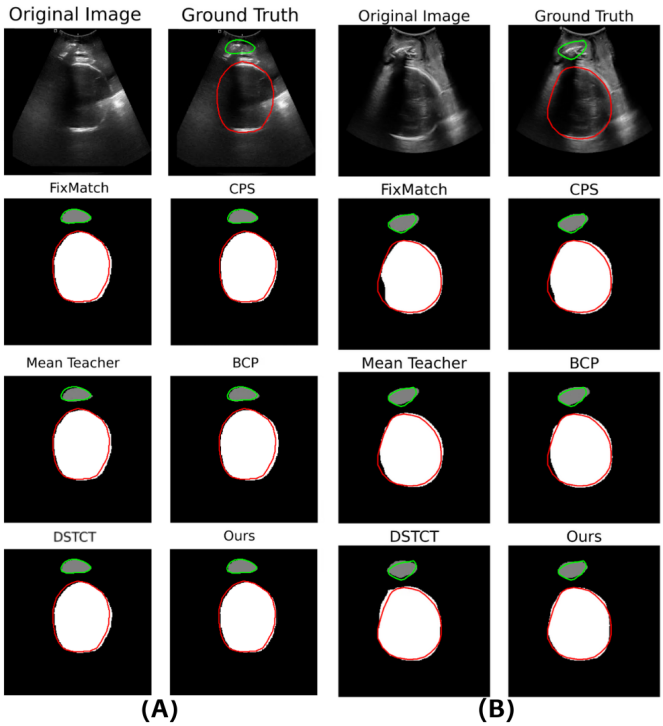

Fig. 4. Qualitative results on the PSFH dataset (A) and external IUGC 2024 dataset (B). Ground-truth contours are shown in green (pubic symphysis) and red (fetal head).

V. DISCUSSION

A. Performance and Results

Our experiments on FUGC and PSFH benchmarks show that OASIS-Net consistently surpasses strong semi-supervised baselines, Mean Teacher, FixMatch, and Cross-Pseudo Supervision (CPS), while using the same DeepLabV3+ backbone and label budget. Concretely, OASIS-Net lifted average Dice on FUGC from 94.3 % (fully-supervised only) to 96.5 %, and reduced HD_{95} from 7.9 px to 3.9 px; similar gains were observed on PSFH, where our dynamic-ramp variant converged 25–35 epochs sooner than the fixed-weight ablation. These improvements indicate that coupling input-space adversarial augmentation (I-FGSM) with weight-space perturbation (DGAP) offers complementary regularization: pixel attacks force the network to cope with speckle-driven appearance shifts, whereas parameter noise smooths sharp minima that commonly arise under scarce supervision.

Ablation confirms this synergy. Adding I-FGSM alone yields a sizable Dice boost (+1.5 pp) relative to the supervised baseline, aligning with prior reports that input perturbations

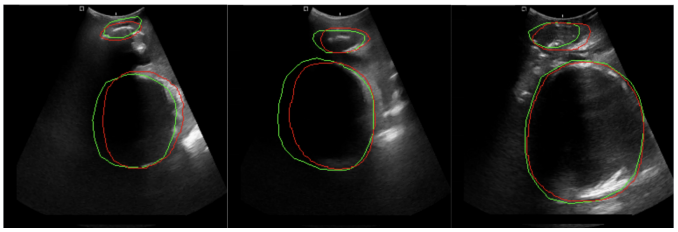

Fig. 5. Some bad qualitative results by *OASIS-Net* with green is GT, red is prediction.

enhance semi-supervised segmentation robustness [19]. Injecting DGAP weight noise on top delivers an additional +0.5 pp Dice and halves HD_{95}, echoing the benefits of model-perturbation frameworks such as GapMatch [9]. Crucially, the final dynamic ramp prevents the over-regularization reported in fixed-alpha studies [20], [21], producing the highest accuracy with the fastest convergence.

Compared with classical consistency schemes, Mean Teacher [10], FixMatch [19], and CPS [16], our dual-space approach yields relative Dice gains of 1–2 pp without extra parameters or test-time overhead. These results complement recent surveys that highlight the need for richer perturbation spaces in medical SSL [22], [23] and demonstrate that a single-network design can match or exceed multi-branch teacher–student frameworks while retaining real-time inference speed. Taken together, the quantitative leap, verified across five-fold FUGC and PSFH cross-validation, provides compelling evidence that dual-space adversarial consistency is a principled, label-efficient route to robust ultrasound segmentation.

Beyond the aggregate metrics, Fig. 5 highlights typical failure cases. In the first two examples, OASIS-Net misidentified the fetal head contour, while in the last case it failed to delineate the inferior pubic symphysis. Such errors closely mirror difficulties faced in manual labor ultrasound: acoustic shadows from the pubic bone and echogenic soft-tissue lines can mimic the fetal skull, and in late-stage labor the inferior symphysis margin often becomes indistinct due to head descent, probe motion, and tissue edema. These observations underscore that even advanced models remain challenged by inherent ultrasound artifacts and anatomical ambiguities, pointing to the need for future work that leverages temporal cues or multimodal guidance.

B. Comparison with Prior Work

Semi-supervised medical segmentation largely follows two lines: (i) pixel-space consistency (e.g., VAT, FixMatch) and (ii) model-ensemble or teacher–student schemes (Mean Teacher, CPS). Consistency methods perturb inputs via intensity jitter, geometric warps, or Mix-style cut–paste and enforce invariant predictions; FixMatch streamlines this with weak/strong augmentations and confidence-filtered pseudo-labels [15]. Yet in noisy ultrasound, such weak perturbations rarely expose worst-case speckle that breaks boundary detection, as noted in recent SSL surveys [24].

Teacher–student methods address label scarcity by letting a slow "teacher" guide a fast "student." Mean Teacher stabilizes training but still depends on input perturbations, while CPS cross-supervises two independently initialized students [16]. Both can inherit confirmation bias when pseudo-masks share systematic errors—an issue magnified in speckled ultrasound.

To mitigate this, adversarial perturbations have been introduced. ASE-Net adds pixel- and image-level discriminators atop self-ensembling [25], and GapMatch perturbs inputs and feature maps without touching weights [9]. Our approach differs in two ways: (1) we use lightweight I-FGSM to synthesize worst-case pixel perturbations on-the-fly, avoiding the adversarial overfit seen in GAN-centric BUS-GAN for breast ultrasound [26]; (2) we pair this with DGAP, a weight-space exploration akin to the diversity mechanism in CMMT-Net [20] but without a dual-head design so OASIS-Net preserves the original parameter count.

Beyond these, shape/anatomy-aware SSL injects priors to prevent implausible masks (e.g., prior-aware pseudo-labeling for ultrasound muscle) [27]. We show that robust dual-space perturbations already suppress spurious structures: ablations indicate HD_{95} drops by 38 % when DGAP is combined with I-FGSM. Finally, concurrent dual-network cervical-ultrasound work reaches similar Dice but requires two ResNet backbones and a contrastive loss [28]; OASIS-Net attains higher Dice with a single backbone and no contrastive branch, underscoring the efficiency of dual-space adversarial consistency.

C. Clinical Relevance and Pregnancy Risks

Accurate segmentation of the cervix and fetal head matters because it underpins time-sensitive obstetric care. A mid-trimester transvaginal cervical length (CL) of ≤ 25 mm defines a short cervix and multiplies the odds of spontaneous preterm birth roughly four- to six-fold [29]; professional guidance therefore recommends universal CL screening at 16-24 weeks and prophylactic vaginal progesterone or selective cerclage when shortening is confirmed [30]. OASIS-Net delineates the anterior and posterior cervical lips, fits the os-to-os (A-to-B) line, and produces automated CL estimates that reduce inter-observer variability and flag cases for clinician review (examples in Fig. 6: 18.7 mm and 19.37 mm). In twin gestations the management evidence is less clear and cerclage is typically reserved for extreme shortening or dilation; automated early-flagging could support timely referral for emergency cerclage, an approach supported by meta-analysis [31].

Cervical Segmentation Challenges: Automated cervical segmentation hinges on clear, diagnostic-quality TVUS frames: a distended bladder can displace or compress the cervix and obscure the internal os, producing ambiguous upper-boundary outlines and biased cervical-length estimates [32]; prominent cervical mucus at the internal os—whose size and echogenicity vary widely—can mimic cervical tissue and confuse landmarking [33]; and placenta previa or a low-lying placenta adjacent to the cervix may present echogenic patterns that are mistaken for cervical boundaries, leading to systematic overestimation of CL with important implications for risk stratification [34]. These acquisition- and contrast-related failure modes are common in practice; pragmatic mitigations such

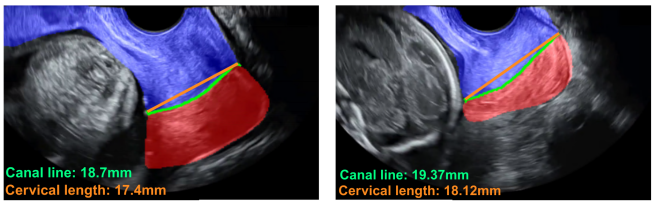

Fig. 6. Automated mid-sagittal cervical-length measurement produced by OASIS-Net. The network segments the anterior (red) and posterior (blue) cervical lips, fits the endocervical canal line (green) connecting the internal os and external os, and reports the resulting length in millimetres, identifying two high-risk pregnancies (≤ 25 mm). Values shown are illustrative using a convenience scaling of $k = 1\,\mathrm{px} : 1\,\mathrm{mm}$; see Methods for pixel-to-mm calibration, sensitivity analysis ($k = 0.8$-$1.5\,\mathrm{mm/px}$), and dataset/ethical compliance.

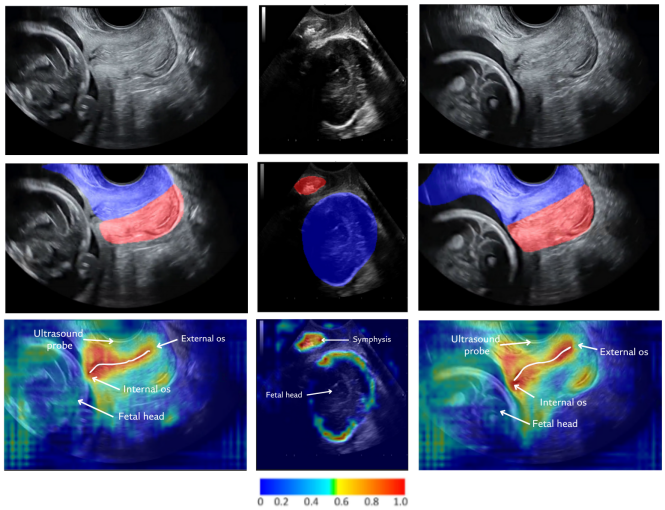

Fig. 7. Interpretability of OASIS-Net: heatmaps on ultrasound frames highlighting internal/external os, probe position, fetal head and pubic symphysis with predicted head contour.

as pre-scan bladder instructions, a brief sonographer quality check, and model uncertainty flags that route ambiguous cases for rapid human review can substantially reduce measurement error and improve the safety of automated pipelines.

Fetal Head Segmentation and Angle of Progression: Beyond prematurity, the network's second target the fetal head pubic symphysis complex, has clear intrapartum value. Automated fetal-head segmentation is routinely used antenatally to compute biometrics such as head circumference and biparietal/partial diameters; during labor, segmenting these landmarks enables objective measurement of the AoP, a validated marker of head descent (AoP $\geq$135° is associated with successful vaginal delivery) [35]. Intrapartum scans are often acquired in busy delivery wards by non-specialist staff and under unstable patient positioning, which increases motion, speckle and shadowing artifacts and makes manual segmentation [36]. Robust, low-latency landmark segmentation and reliable automated AoP estimation at the bedside can therefore reduce inter-operator variability, speed triage, and support timely decisions about mode of delivery. We emphasize that this AI model is intended as a clinical decision-support tool to assist transperineal ultrasound in labor by providing more objective, consistent, and rapid measurements to inform clinicians; it is not a substitute for expert assessment, and final management decisions should remain with the care team.

Clinical Applications: Finally, automated delineation unlocks downstream analytics: serial head measurements support early detection of growth disorders, and precise cervix masks enable biomechanical modeling of tissue remodeling, each application benefiting from label-efficient training on routinely archived, yet unlabeled, clinical scans.

D. Limitations

Despite encouraging results, *OASIS-Net* entails practical trade-offs. First, dual-space adversarial training introduces extra hyper-parameters (attack radius, iteration count K) that materially influence convergence; even modest changes can shift Dice by 2–3 points on small validation sets [25], [37]. Automating their calibration remains open. Second, computation is heavier than classic consistency learning: I-FGSM adds up to K forward–backward passes per unlabeled frame, and DGAP incurs an additional weight-space step. Similar memory ceilings are reported for 3-D adversarial defenses and "Free" adversarial training [38]. Third, robustness under domain shift and broader generalization are not guaranteed. Differences in probe frequency and patient habitus can degrade cross-scanner transfer (Dice drops of 8–12 points without adaptation) [39]. While adversarial weight noise helps marginally, *OASIS-Net* would benefit from domain alignment or test-time adaptation [40] and prospective multi-center validation on larger benchmarks (e.g., SSL4MIS) [24]. Finally, a clinical caveat: AI-assisted transperineal measurement of the AOP is reproducible, especially in suspected macrosomia, malpresentation, or late first-stage delayed descent, and our PS–FH dataset/algorithm improves fidelity over standard ultrasound; however, AOP should remain an adjunct tool, as digital examination is still essential for assessing labor readiness.

VI. Conclusion and Future Work

In this paper we introduce OASIS-Net, a semi-supervised segmentation framework that enforces a unified consistency loss across input-level adversarial attacks and gradient-aligned weight perturbations. Using a single DeepLabV3+ backbone, OASIS-Net achieves state-of-the-art performance on two obstetric ultrasound benchmarks while substantially reducing labeled-data requirements; ablations confirm the complementary benefits of pixel- and parameter-space perturbations and the dynamic ramp. Clinically, the model produces high-precision outputs for automated cervical length. Future work will extend OASIS-Net to 3-D/cine ultrasound, pursue cross-center domain adaptation, and develop lightweight model for real-time deployment on portable devices.

Acknowledgments

The authors thank the PSFH, FUGC, and IUGC Challenge organizers for the datasets [12]–[14]. Fig. 2 is reproduced with permission under a paid license from Dr. Than Trong Thien. This work was supported by the National Science and Technology Council, Taiwan [grant numbers MOST111-2628-E-038-002-MY3 and NSTC114-2221-E-038-015].

References

[1] J. Owen, J. M. Szychowski, G. Hankins, J. D. Iams, J. S. Sheffield, A. Perez-Delboy, V. Berghella, D. A. Wing, E. R. Guzman, V. U. T. Consortium *et al.*, "Does midtrimester cervical length 25 mm predict preterm birth in high-risk women?" *American journal of obstetrics and gynecology*, vol. 203, no. 4, pp. 393–e1, 2010.

[2] N. Baños, X. P. Burgos-Artizzu, B. Valenzuela-Alcaraz, D. Coronado-Gutiérrez, Á. Perez-Moreno, J. Ponce, E. Gratacós, and M. Palacio, "Intra-and interobserver reproducibility of second trimester ultrasound cervical length measurement in a general population," *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 35, no. 5, pp. 999–1002, 2022.

[3] J. E. Jung and Y. J. Lee, "Intrapartum transperineal ultrasound: angle of progression to evaluate and predict the mode of delivery and labor progression," *Obstetrics & Gynecology Science*, vol. 67, no. 1, pp. 1–16, 2024.

[4] S. Guo, Z. Liu, Z. Yang, C. H. Lee, Q. Lv, and L. Shen, "Multi-scale multi-object semi-supervised consistency learning for ultrasound image segmentation," *Neural Networks*, vol. 184, p. 107095, 2025.

[5] S. M. A. Naqvi, M. Shabaz, M. A. Khan, and S. I. Hassan, "Adversarial attacks on visual objects using the fast gradient sign method," *Journal of Grid Computing*, vol. 21, no. 4, p. 52, 2023.

[6] B. Zhang, Y. Wang, Y. Hou, and et al., "Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling," in *NeurIPS*, 2021.

[7] T. Miyato, S.-i. Maeda, M. Koyama, and S. Ishii, "Virtual adversarial training: a regularization method for supervised and semi-supervised learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1979–1993, 2018.

[8] Z. Wei, J. Zhu, and Y. Zhang, "Sharpness-aware minimization alone can improve adversarial robustness," *arXiv preprint arXiv:2305.05392*, 2023.

[9] W. Huang, L. Zhang, Z. Wang, and Y. Wang, "Gapmatch: Bridging instance and model perturbations for enhanced semi-supervised medical image segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 4. AAAI, 2024, pp. 4772–4780. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/28351

[10] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *arXiv preprint arXiv:1703.01780*, 2017.

[11] Y. Zhu, X. Wang, T. Liu, and Y. Fu, "Multi-perspective dynamic consistency learning for semi-supervised medical image segmentation," *Scientific Reports*, vol. 15, no. 1, pp. 1–15, 2025.

[12] J. Bai, Z. Yang, M. K. Hasan, J. Gan, Z. Liang, W. Cai, T. Tan, J. Ye, M. Yaqub, D. Ni, S. Slimani, B. Ohene-Botwe, R. V. M. Campello, and K. Lekadir, "Fetal ultrasound grand challenge: Semi-supervised cervical segmentation (fugc25)," ISBI 2025 Dataset, 2024. [Online]. Available: https://zenodo.org/records/14328192

[13] J. Bai, Z. Zhou, Z. Ou, G. Koehler, R. Stock, K. Maier-Hein, M. Elbatel, R. Martí, X. Li, Y. Qiu *et al.*, "Psfhs challenge report: pubic symphysis and fetal head segmentation from intrapartum ultrasound images," *Medical Image Analysis*, vol. 99, p. 103353, 2025.

[14] J. Bai, K. Lekadir, D. Ni, S. Slimani, V. Campello, B. Ohene-Botwe, Y. Lu, G. W. Chen, H. Hou, D. Qiu, and Z. Zhou, "Intrapartum ultrasound grand challenge 2024 (miccai2024)," 12 2024.

[15] K. Sohn, D. Berthelot, N. Carlini *et al.*, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," in *NeurIPS*, 2020.

[16] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *CVPR*, 2021.

[17] Y. Bai, D. Chen, Q. Li, W. Shen, and Y. Wang, "Bidirectional copy-paste for semi-supervised medical image segmentation," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 11 514–11 524.

[18] H.-H. Pham, L. Tran Quoc Khanh, H.-T. Nguyen, N. L. V. Vu, Q.-V. Dinh, and T.-H. Nguyen, "Fetal-bcp: Addressing empirical distribution gap in semi-supervised fetal ultrasound segmentation," in *IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2024.

[19] P. Upretee and B. Khanal, "Fixmatchseg: Fixing fixmatch for semi-supervised semantic segmentation," *arXiv preprint arXiv:2208.00400*, 2022.

[20] W. Li *et al.*, "Diversity matters: Cross-head mutual mean-teaching for semi-supervised medical image segmentation," *Medical Image Analysis*, 2024.

[21] J. Zhu, B. Bolsterlee, B. V. Chow, Y. Song, and E. Meijering, "Hybrid dual mean-teacher network with double-uncertainty guidance for semi-supervised segmentation of magnetic resonance images," *Computerized Medical Imaging and Graphics*, vol. 115, p. 102383, 2024.

[22] J. Guo, Z. Li, and Y. Lin, "Semi-supervised learning for real-time segmentation of ultrasound video objects: A review." *Advanced Ultrasound in Diagnosis & Therapy (AUDT)*, vol. 7, no. 4, 2023.

[23] A. A. of Family Physicians (AAFP), "Cervical length screening for preterm birth," *Am Fam Physician*, vol. 107, no. 3, pp. 174–180, 2023, available at: https://www.aafp.org.

[24] M. Kots, M. Pozigun, A. Konstantinov, and V. Chukanov, "Semi-supervised learning for medical image segmentation," in *Proceedings of International Scientific Conference on Telecommunications, Computing and Control: TELECCON 2019*. Springer, 2021, pp. 245–253.

[25] T. Lei *et al.*, "Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network," *IEEE Transactions on Medical Imaging*, 2023.

[26] H. Xie, C. Fu, X. Zheng, Y. Zheng, C.-W. Sham, and X. Wang, "Adversarial co-training for semantic segmentation over medical images," *Computers in biology and medicine*, vol. 157, p. 106736, 2023.

[27] Y. Chen, Y. Wang, Z. Zheng, J. Hu, Y. Shi, S. Xiong, X. X. Zhu, and L. Mou, "Striving for simplicity: Simple yet effective prior-aware pseudo-labeling for semi-supervised ultrasound image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 604–614.

[28] F. Wang, K. M. Curran, and G. Silvestre, "Semi-supervised cervical segmentation on ultrasound by a dual framework for neural networks," in *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2025, pp. 1–5.

[29] J. Gilner and J. Biggio, "Management of short cervix during pregnancy: a review," *American journal of perinatology*, vol. 33, no. 03, pp. 245–252, 2016.

[30] S. for Maternal-Fetal Medicine (SMFM), "Smfm consult series no. 60: The role of transvaginal cervical length screening for preterm birth prevention," *American Journal of Obstetrics and Gynecology*, vol. 228, no. 5, pp. B2–B12, 2023.

[31] C. Chatzakis, A. Efthymiou, A. Sotiriadis, and G. Makrydimas, "Emergency cerclage in singleton pregnancies with painless cervical dilatation: A meta-analysis," *Acta obstetricia et gynecologica Scandinavica*, vol. 99, no. 11, pp. 1444–1457, 2020.

[32] C. M. Coutinho, A. Sotiriadis, A. Odibo, A. Khalil, F. D'antonio, H. Feltovich, L. Salomon, P. Sheehan, R. Napolitano, V. Berghella *et al.*, "Isuog practice guidelines: role of ultrasound in the prediction of spontaneous preterm birth," *Ultrasound in obstetrics & gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, vol. 60, no. 3, pp. 435–456, 2022.

[33] J. T. Thomas, S. G. Petersen, J. Chua, S. Connard, K. Gibbons, R. Cincotta, and G. Gardener, "Prominent cervical mucous and its impact on cervical length measurement: Findings of a national survey," *Australian and New Zealand Journal of Obstetrics and Gynaecology*, vol. 54, no. 2, pp. 108–116, 2014.

[34] A. Altraigey, M. Ellaithy, E. Barakat, and A. Majeed, "Cervical length should be measured for women with placenta previa: cohort study," *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 34, no. 13, pp. 2124–2131, 2021.

[35] S. Namouz, S. Porat, N. Okun, R. Windrim, and D. Farine, "Emergency cerclage: literature review," *Obstetrical & gynecological survey*, vol. 68, no. 5, pp. 379–388, 2013.

[36] T. Ghi, T. Eggebø, C. Lees, K. Kalache, P. Rozenberg, A. Youssef, L. Salomon, and B. Tutschek, "Isuog practice guidelines: intrapartum ultrasound," *Ultrasound in Obstetrics & Gynecology*, vol. 52, no. 1, pp. 128–139, 2018.

[37] X. Zhao and W. Wang, "Semi-supervised medical image segmentation based on deep consistent collaborative learning," *Journal of Imaging*, vol. 10, no. 5, p. 118, 2024.

[38] M. Shiri, A. Bruno, and D. Loiacono, "Memory-efficient 3d high-resolution medical image synthesis using crf-guided gans," in *International Conference on Pattern Recognition*. Springer, 2024, pp. 184–194.

[39] E. Chen, T. S. Mathai, V. Sarode, H. Choset, and J. Galeotti, "A study of domain generalization on ultrasound-based multi-class segmentation of arteries, veins, ligaments, and nerves using transfer learning," *arXiv preprint arXiv:2011.07019*, 2020.

[40] H. Guan and M. Liu, "Domain adaptation for medical image analysis: a survey," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 3, pp. 1173–1185, 2021.