Supplementary Material — Part B: Implementation Details & Visualizations

This document provides: (B1) additional implementation details; (B2) extended experimental results; (B3) qualitative visualizations.

B1. Additional Implementation Details

B1.1. Structure Tensor Computation

The STP module computes the structure tensor at four fixed scales $\sigma_k \in \{1.0, 2.0, 3.0, 4.0\}$ pixels, one per decoder level. The image gradient is computed via Sobel operators:

$$K_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad K_y = K_x^\top. \tag{B1}$$

The Gaussian integration kernel G_{σ_k} is a full 2D Gaussian convolution (not separable 1D) for accurate isotropy at all scales. All operations run under `torch.no_grad()` in `float32`.

B1.2. ADGM: Decoupled Edge Gate and Direction MLP

The Edge Gate and Direction MLP are **fully decoupled** modules:

- **Edge Gate** — operates on the feature map F: a 1×1 convolution followed by sigmoid, producing per-pixel spatial attention $\alpha \in (0, 1)$.
- **Direction MLP** — operates on geometric inputs (e_{1x}, e_{1y}) from the STP descriptor only:

$$\text{Dir-MLP}: \mathbb{R}^2 \to \mathbb{R}^1, \quad [2,\ 16,\ 1], \quad \text{ReLU + Sigmoid}, \tag{B2}$$

  producing direction weight $\beta \in (0, 1)$.

The two branches are fused only at the final gated residual step. The Direction MLP has $2{\times}16{+}16{+}16{\times}1{+}1 = 49$ parameters per decoder level, totaling **196 parameters** across all four levels — negligible overhead.

B1.3. Loss Hyperparameters and Ramp-Up Schedule

The geometric loss weight follows a Gaussian ramp-up:

$$\lambda_{\text{geo}}(t) = \lambda_{\max} \cdot \exp\left(-5\left(1 - \frac{t}{T_{\text{ramp}}}\right)^2\right), \tag{B3}$$

with $\lambda_{\max} = 0.05$ and $T_{\text{ramp}} = 15$ epochs. This configuration achieves the best boundary accuracy on the HC18 validation set: geometric refinement activates rapidly within the first 15 epochs while semantic learning stabilizes. Full loss weights: $\alpha_{\text{curv}} = 2.0$, $\beta_{\text{struct}} = 0.5$, $\kappa_0 = 0.5$, $\tau_{\text{curv}} = 10$, $\varepsilon = 10^{-6}$, $\sigma_p = 1.0$ (curvature pre-smoothing).

B2. Qualitative Visualizations

B2.1. Predictive Epistemic Uncertainty via TTA

We construct epistemic uncertainty heatmaps via Test-Time Augmentation (TTA): $N{=}8$ spatial transforms (horizontal/vertical flip, orthogonal rotations) yield an ensemble of probability maps; pixel-wise variance is the uncertainty estimate.

As shown in Fig. B1, STGA-Net shows high confidence (low variance, cold colors) over uniform background and lesion cores, while uncertainty (warm/red) concentrates strictly along ambiguous tumor margins and acoustic shadows. This anatomically meaningful uncertainty distribution confirms that $\mathcal{L}_{\text{CSAB}}$-guided training localizes uncertainty at geometrically challenging transitions rather than spreading it uniformly.

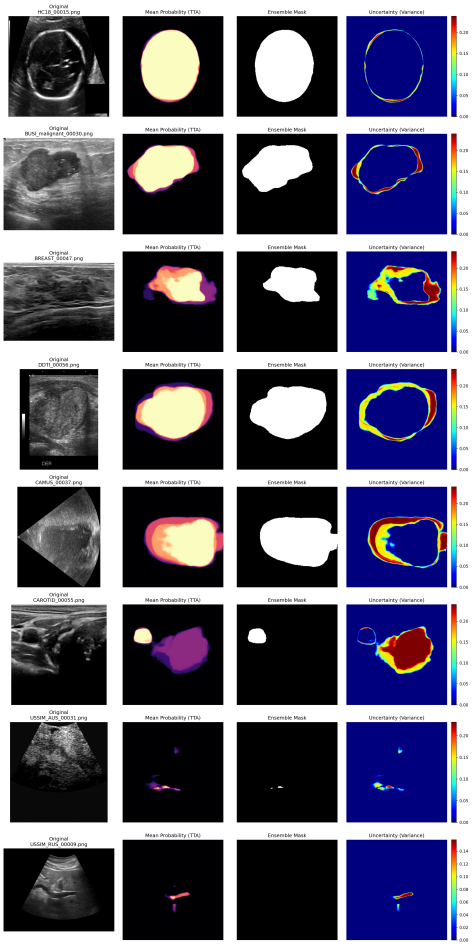

Figure B1. **Epistemic uncertainty via TTA (Fig. B1).** Each row (one dataset): original B-mode image | mean probability map over $N{=}8$ passes | binarized mask | variance heatmap (jet, 0–0.20). Rows top-to-bottom: HC18, BUSI, BrEaST, DDTI, CAMUS, Carotid, USSIM_AUS, USSIM_RUS.

B2.2. PVTv2 Stage-4 Feature Activation Mapping

We extract Stage-4 PVTv2 features [2] and compute the mean absolute activation across channels to produce an activation heatmap. Fig. B2 shows that activation peaks align with anatomical target structures across all six domains without per-pixel supervision. Notably, the transformer encoder captures not only lesion interiors but also surrounding fascial planes and acoustic shadows, enabling discrimination of genuine pathology from speckle artifacts.

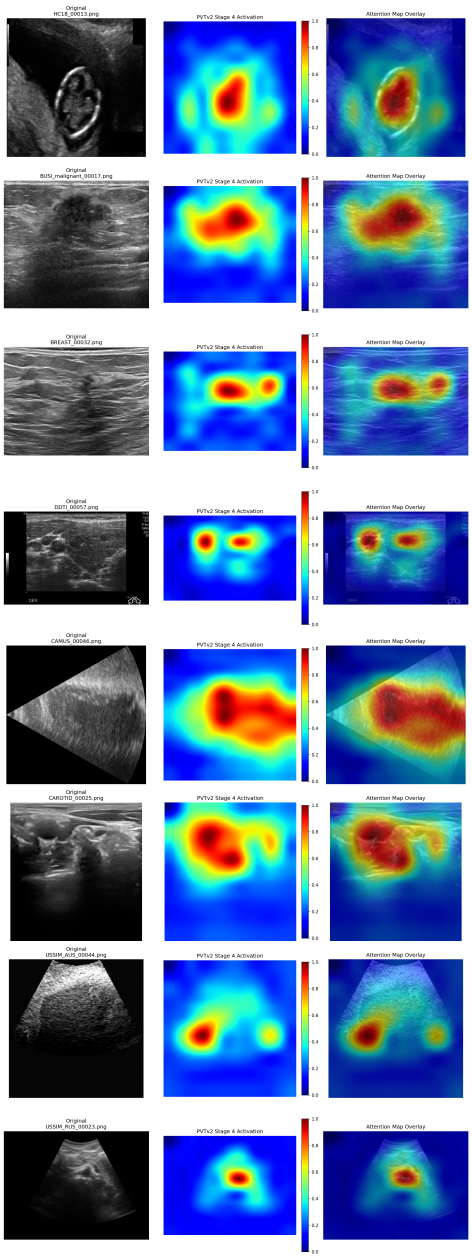

Figure B2. **PVTv2 Stage-4 activation overlay (Fig. B2).** Each sample: original image | activation heatmap (jet, 0–1) | overlay on scan. Rows top-to-bottom: HC18, BUSI, BrEaST, DDTI, CAMUS, Carotid, USSIM_AUS, USSIM_RUS.

B2.3. Student vs. EMA Teacher Pseudo-Mask Comparison

Fig. B3 compares student predictions with EMA teacher pseudo-masks at the best checkpoint across six anatomical domains. The teacher, updated via exponential moving average [1], consistently yields tighter, more coherent boundaries, with reduced fragmentation on challenging cases

(BUSI multi-focal lesions, irregular abdominal structures). This confirms that EMA smoothing combined with $\mathcal{L}_{\text{CSAB}}$ geometric regularization suppresses confirmation bias in pseudo-label generation.

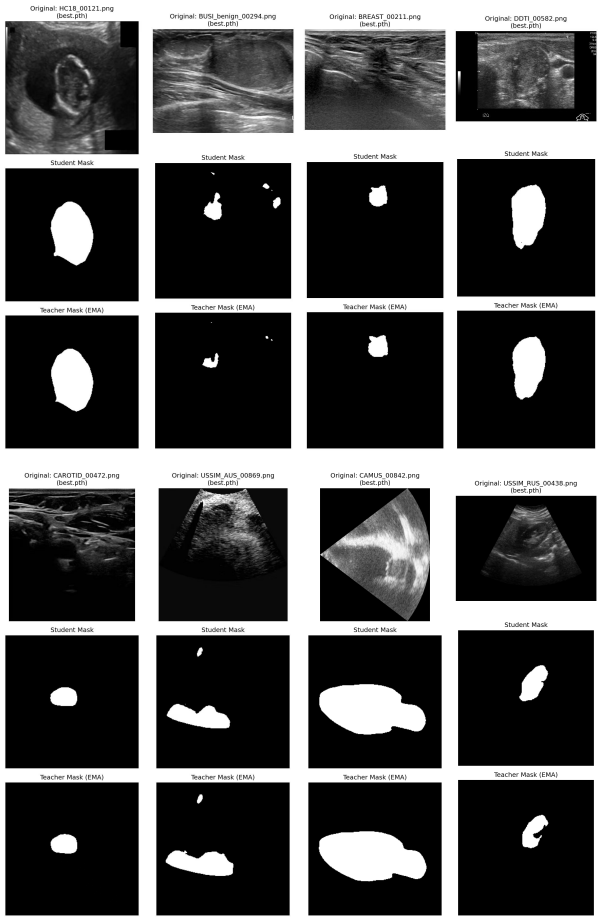

Figure B3. **Student vs. EMA teacher pseudo-masks (Fig. B3).** Each panel: original image (top) | student mask (mid) | teacher pseudo-mask (bottom). *Top row:* HC18, BUSI, BrEaST, DDTI. *Bottom row:* Carotid, USSIM_AUS, CAMUS, USSIM_RUS.

References

[1] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. 2

[2] Wenhai Wang, Enze Xie, Xiang Li, et al. PVTv2: Improved baselines with Pyramid Vision Transformer. *Computational Visual Media*, 8(3):415–424, 2022. 2