SonoSPADE: Supplementary Material

Thanapong Intharah[1], Zhichao Gao[2], and Hao Dong[2]

[1] Visual Intelligence Laboratory, Khon Kaen University, Khon Kaen, Thailand
thanapong.intharah@gmail.com
[2] School of Computer Science, Peking University, Beijing, China

This supplement collects the whole-frame FID/KID scores (Table S1), the training-term ablation (Table S2), the per-tissue label-swap locality breakdown (Table S3), the per-tissue qualitative result (Fig. S1), the pseudo-label quality of the frozen S_{US} on real frames (Table S4), the complete hyperparameters (Table S5) referenced in the main paper. All numbers are generated from the same result files as the main tables. For the head-to-head, all three translators are trained on the *same* real pool; an off-pool CUT checkpoint wrongly reverses the result.

Table S1. Whole-frame FID/KID ($\times 10^3$) against the real pool (InceptionV3), lower better. SonoSPADE is not competitive by design, because these metrics reward the global recolor it does not perform. The bottom block re-runs in the sector geometry, which improves all absolute scores but not the ranking.

Method	FID ↓	KID $\times 10^3$ ↓
Raw physics	315.3	294.3
CycleGAN	**192.2**	203.1
CUT	235.6	245.9
S-CycleGAN (reimpl.)	199.7	**196.9**
SonoSPADE	294.7	306.5
Raw physics-curvi	212.9	265.4
CycleGAN-curvi	**67.9**	**55.6**
CUT-curvi	72.3	55.9
S-CycleGAN-curvi	104.2	84.2
SonoSPADE-curvi	191.1	232.6

Table S2. Ablation (one term off per row, each retrained at a fixed budget). Removing the frozen S_{US} seg-consistency term roughly halves locality and degrades structure, the largest single effect. This is the full version of the ablation summarized in Sec. 4.5 of the main paper.

Variant	Edge IoU ↑	Organ W_1 ↓	Locality ↑
Full SonoSPADE	0.41	0.115	27.1
− S_{US} consistency	0.30	0.129	13.7
− OASIS LabelMix	0.39	0.111	28.6
− temporal consistency	0.42	0.106	26.3

Table S3. Per-tissue label-swap locality on the physics/label content frames (higher means the texture change is more localized to the relabeled region). Liver is the reference tissue and is excluded; spleen and muscle rarely reach the 200-pixel minimum in the content frames and are omitted. The values are the measured per-class means behind the main-paper locality of 26.3 (they average to it). Per-frame standard deviations can be added with `scripts/eval_locality_std.py`.

Tissue	Label-swap locality ↑
Gallbladder	22.8
Vessel	18.7
Kidney	30.2
GI tract	40.3
Pancreas	28.8
Fat	10.4
Bone	60.7
Lung	13.0
Soft tissue	11.5
All tissues (mean)	26.3
Label-free translators	0 (by construction)

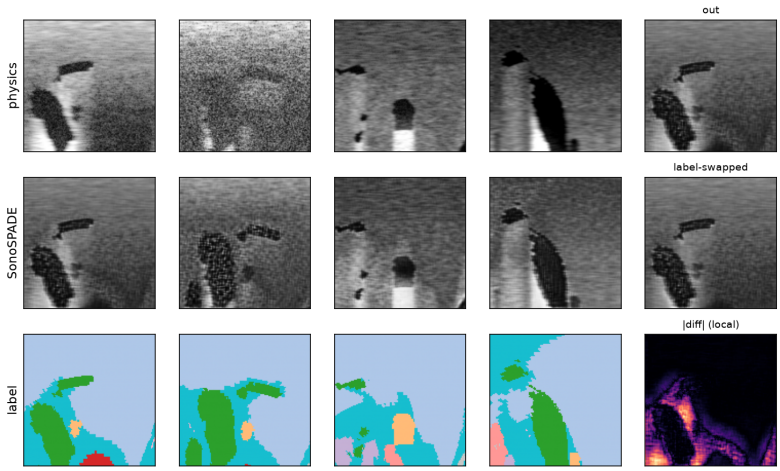

Fig. S1. Per-tissue qualitative result (Sec. 4.2 of the main paper). The rows are the physics render, the SonoSPADE output, and the tissue label. Each tissue gets its own speckle, and relabeling a region changes its texture locally (last column): the swapped region's texture changes while the rest of the frame stays put, which is what the label-swap locality metric measures.

Table S4. Pseudo-label quality of the frozen S_{US} on the 60 annotated real Kaggle frames, which it never saw (Sec. 4.3 of the main paper). Linear is the checkpoint that produced the reported pseudo-labels; curvilinear is the geometry-matched retrain. The reference masks annotate only these five organs, so any prediction on unannotated tissue counts against precision, which is why precision, not sensitivity, is the binding term. Trained only in simulation, S_{US} gives the discriminator a coarse regional prior rather than accurate anatomy.

		Linear			Curvilinear		
Organ	Frames	Dice	Sens.	Prec.	Dice	Sens.	Prec.
Liver	39	0.124	0.23	0.09	0.339	0.38	0.31
Gallbladder	12	0.000	0.00	0.00	0.030	0.04	0.03
Vessel	16	0.017	0.40	0.01	0.058	0.30	0.03
Kidney	15	0.000	0.00	0.00	0.000	0.00	0.00
Spleen	6	0.000	0.00	0.00	0.000	0.00	0.00
Macro	60	0.028			0.085		

Table S5. Complete hyperparameters for the reported SonoSPADE run (Sec. 3 of the main paper), as used by `scripts/run_sonospade_experiment_real.sh`. Loss weights follow $\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{nce}\mathcal{L}_{nce} + \lambda_{seg}\mathcal{L}_{seg} + \lambda_{tc}\mathcal{L}_{tc}$ and $\mathcal{L}_D = \mathcal{L}^D_{adv} + \lambda_{lm}\mathcal{L}_{lm} + \frac{\gamma}{2}\mathcal{L}_{R1}$.

Group	Setting	Value
Loss weights	λ_{seg} (frozen S_{US} consistency)	5
	λ_{nce} (PatchNCE content)	1
	λ_{tc} (temporal consistency)	1
	λ_{lm} (OASIS LabelMix, on D)	1
	γ (R1 penalty, on D)	1
	Dice share inside $\mathcal{L}_{seg}$	1 (unweighted CE+Dice)
Optimizer	Adam, learning rate	2×10^{-4}
	(β_1, β_2)	$(0.5, 0.999)$
	Batch size	4
	Iterations	1500
	Training resolution	128×128
	Seed	0
Generator	Base width, SPADE blocks	32, 6
	SPADE hidden width	64
	Canonical classes K	13
	EMA decay (deployed weights)	0.999
Discriminator	U-Net depth, base width	4, 32
	Per-pixel output classes	10 (N=9 grouped + fake)
PatchNCE	Patches per layer	256
	Temperature	0.07