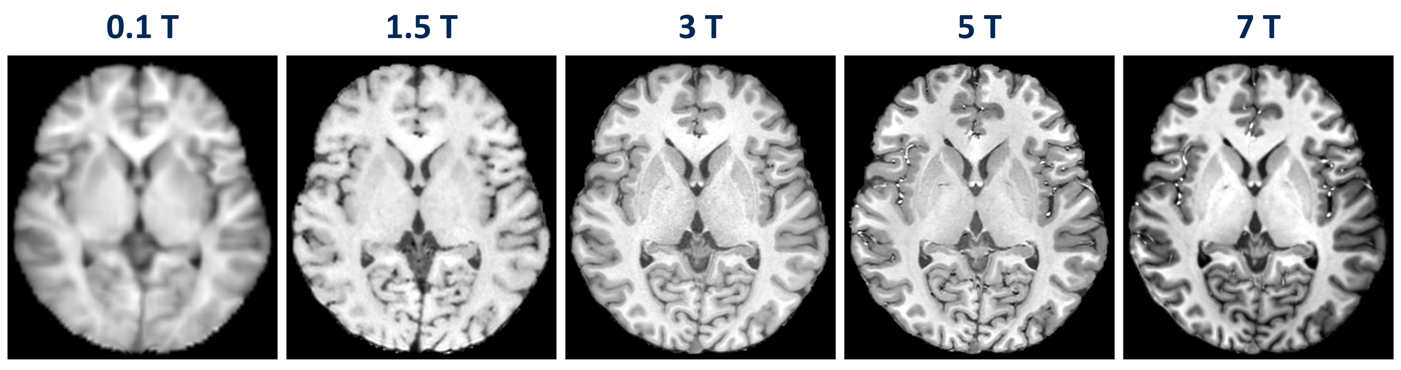
Cross-Field MRI Synthesis with an Unpaired Neural Schrödinger Bridge

Titouan Le Gourrierec · **HyoSeok Lee** · **Sung-Hong Park** Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology · titouanlegourrierec@kaist.ac.kr

1 Task definition

Synthesise the 7 T-equivalent volume from an input at 0.1 / 1.5 / 3 / 5 T, in any of three contrasts (T1w, T2w, T2-FLAIR) — up to 12 separate models allowed. Only 3 paired subjects (660 slices) sit in the training set.



0.1–7 T · 5 fields, 3 vendors | **2,300+** cases · 1,900+ unpaired | **40** travelling volunteers | **660** paired training slices

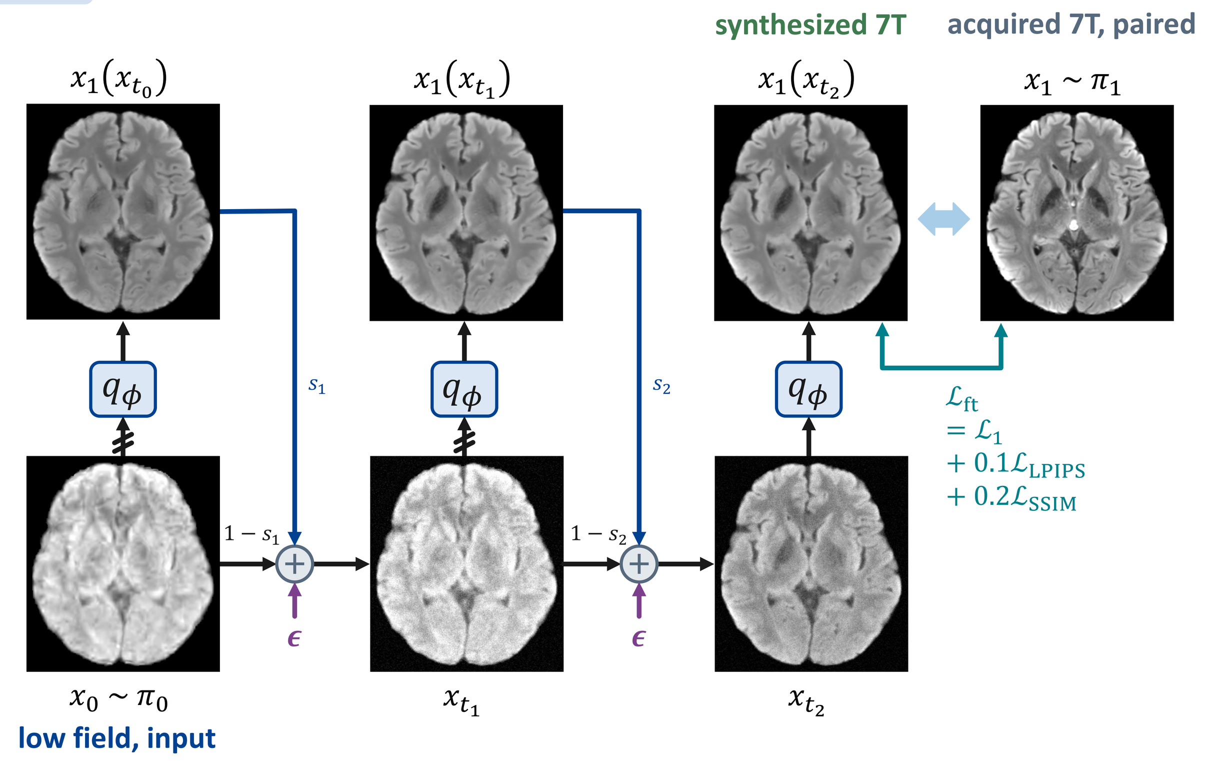
The same travelling volunteer at five fields (T1w shown) — SNR, resolution, contrast and homogeneity all shift with B_0, differently per contrast.

Paired supervision is scarce → the field map must be learned from unpaired data.
Ranking: Σ of per-metric ranks across 36 teams — nRMSE ↓ · LPIPS ↓ · SSIM ↑ · Dice ↑ · Vol. consistency ↑ (lowest Σ wins).

2 Method overview

- **Neural Schrödinger bridge formulation**
  An entropy-regularised Schrödinger Bridge [1]: 3 refinement steps of one shared network q_ϕ — with CUT's PatchNCE + identity terms [2] protecting anatomy.
- **Spatial and cross-contrast context**
  2.5D neighbours + the two co-registered contrasts, injected by first-convolution inflation — the same function at initialisation.
- **Multi-field, multi-contrast training**
  Train 0.1→7 T from scratch, warm-start 1.5 / 3 / 5 T (+5 epochs each): **201 GPU-h instead of ≈ 27 GPU-days.**
- **Intensity calibration at inference**
  Two global scalars (mean + residual LS scale): **−12.1 % nRMSE, −6.7 % LPIPS** — for free.

2.1 Unpaired Neural Schrödinger Bridge



SAMPLING SCHEME

$$x_{t_{i+1}} = (1-s_{i+1})x_{t_i} + s_{i+1}\mathrm{sg}[x_1(x_{t_i})] + \epsilon$$

$$s_{i+1} \coloneqq (t_{i+1}-t_i)/(1-t_i)$$

$$\sigma_{i+1} = s_{i+1}(1-s_{i+1})(1-t_i)\tau$$

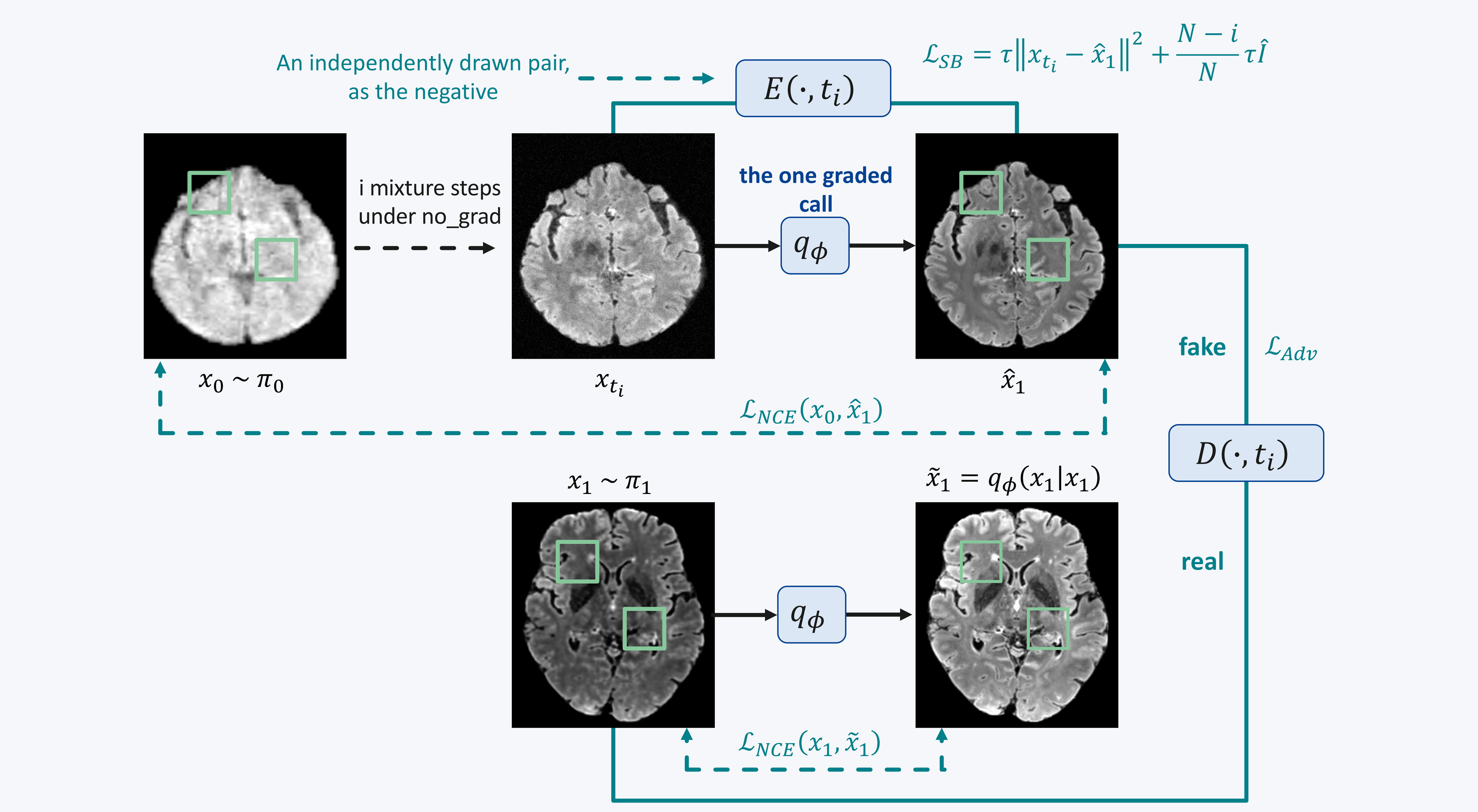
$$\epsilon \sim \mathcal{N}(0, \sigma_{i+1} I)$$

$N = 3 \quad \tau = 0.01$

0 — 0.5 — 0.83 — 1 (t_0, t_1, t_2, t_3)

A one-step translator (GAN [4], CycleGAN [3], CUT [2]) suits domains that differ mainly in texture — and may not explicitly constrain anatomical content or assumes an invertibility a low-field scan cannot satisfy. Across fields, SNR, resolution, contrast and artefacts shift at once, so **we keep CUT's content terms and replace its single generator step with an optimal-transport bridge [1]**: a Markov chain over 0 = t_0 < … < t_n = 1, one time-conditional network q_ϕ, each step refining the predicted 7 T.

STAGE 1 — LEARN THE FIELD MAP FROM UNPAIRED DATA



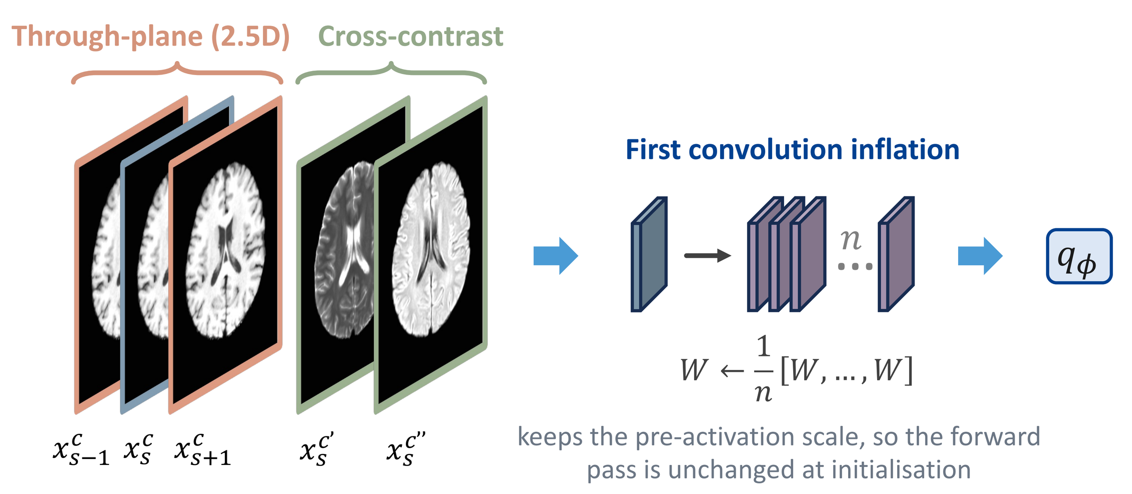
λ_SB = λ_GAN = λ_NCE = 1 · identity $q_\phi(x_1) \approx x_1$ · N = 3 steps, τ = 0.01 — one graded pass of q_ϕ · E, D discarded after this stage

$$\mathcal{L}_{UNSB}(\phi, t_i) = \lambda_{GAN}\mathcal{L}_{Adv} + \lambda_{SB}\mathcal{L}_{SB} + \lambda_{NCE}/2\,[\mathcal{L}_{NCE}(x_0, \hat{x}_1) + \mathcal{L}_{NCE}(x_1, \tilde{x}_1)]$$

STAGE 2 — FINE-TUNE ON THE 660 PAIRED SLICES

The same network and inference procedure are used. $\mathcal{L}_{\mathrm{ft}} = \mathcal{L}_1 + 0.1\mathcal{L}_{\mathrm{LPIPS}} + 0.2\mathcal{L}_{\mathrm{SSIM}}$

2.2 Input representation

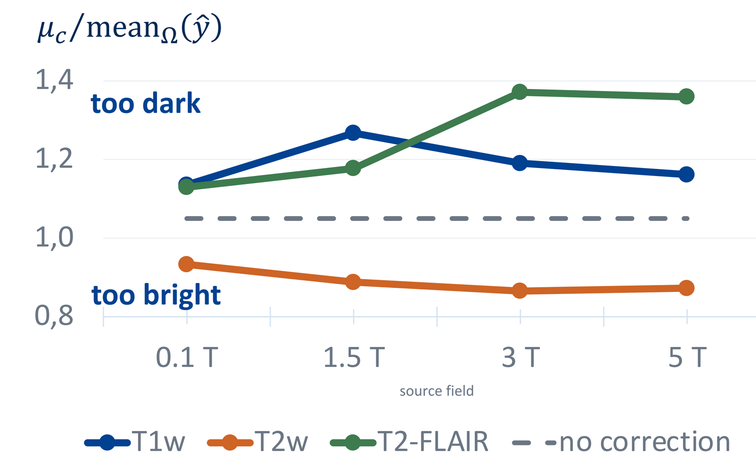


keeps the pre-activation scale, so the forward pass is unchanged at initialisation

	2.5D	Cross-contrast		
		T1	T2	T2 FLAIR
0.1 T	✓	✓	✓	✓
1.5 T	✓	✓	✓	✓
3 T	✓	×	×	✓
5 T	✓	×	×	✓

*The contrasts are already co-registered, so no additional registration is required

2.3 Global intensity calibration



INTENSITY CALIBRATION

$$\hat{y} \leftarrow \mathrm{clip}\left(\frac{\mu_c}{\mathrm{mean}_\Omega(\hat{y})} \cdot \hat{y}, 0, 1\right)$$

$$\Omega = (x > \epsilon) \cap (\hat{y} > \epsilon)$$

μ is the median brain mean over **training** 7 T volumes — 0.1867 / 0.2183 / 0.2170 for T1w / T2w / T2-FLAIR.

−12.1 % nRMSE | **−6.7 %** LPIPS | **+1.1 %** SSIM

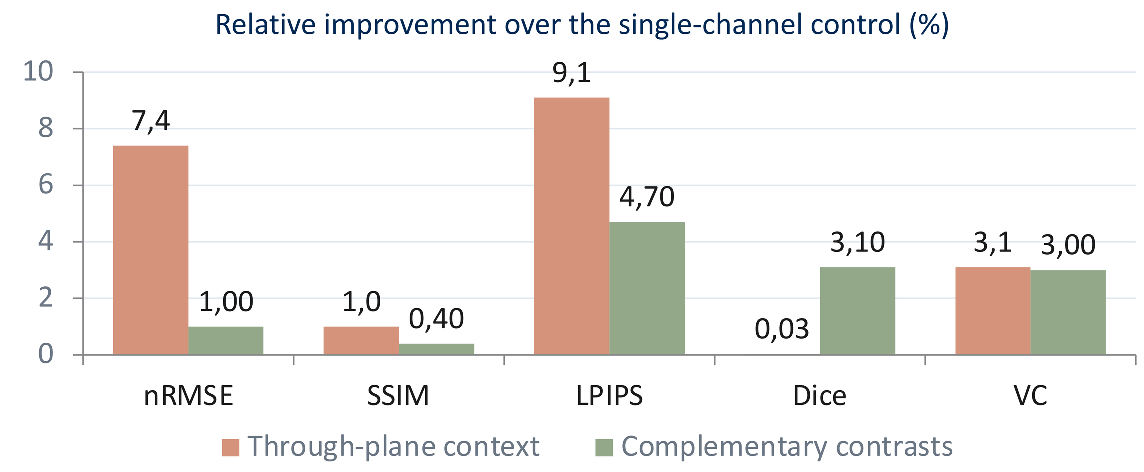
3.1 Validation results

Ranked **1st/36** according to the challenge composite ranking on the validation leaderboard (Σ = 13). The method achieved the highest Dice and volume-consistency scores, while ranking 3rd–4th on image-quality metrics.

#	Method	nRMSE ↓	LPIPS ↓	SSIM ↑	Dice ↑	VC ↑	Σ ↓
1	**Ours**	**0.2595 (4)**	**0.0676 (3)**	**0.9244 (4)**	**0.9071 (1)**	**0.9168 (1)**	**13**
2	Team A	0.2643 (5)	0.0637 (2)	0.9259 (3)	0.9064 (2)	0.9097 (3)	15
3	Team B	**0.2591 (2)**	**0.0587 (1)**	**0.9293 (1)**	0.8884 (7)	0.8828 (7)	18
4	Team C	0.2775 (6)	0.0704 (4)	0.9188 (6)	0.8890 (6)	0.8972 (4)	26
5	Team D	0.2880 (10)	0.0721 (5)	0.9137 (7)	0.9002 (4)	0.8869 (6)	32
—	CUT baseline	0.3041 (30)	0.0750 (10)	0.9053 (19)	0.8543 (14)	0.8601 (17)	90

Best shown value per metric in bold navy · per-metric rank in parentheses · Σ = sum of the five ranks (lower is better).

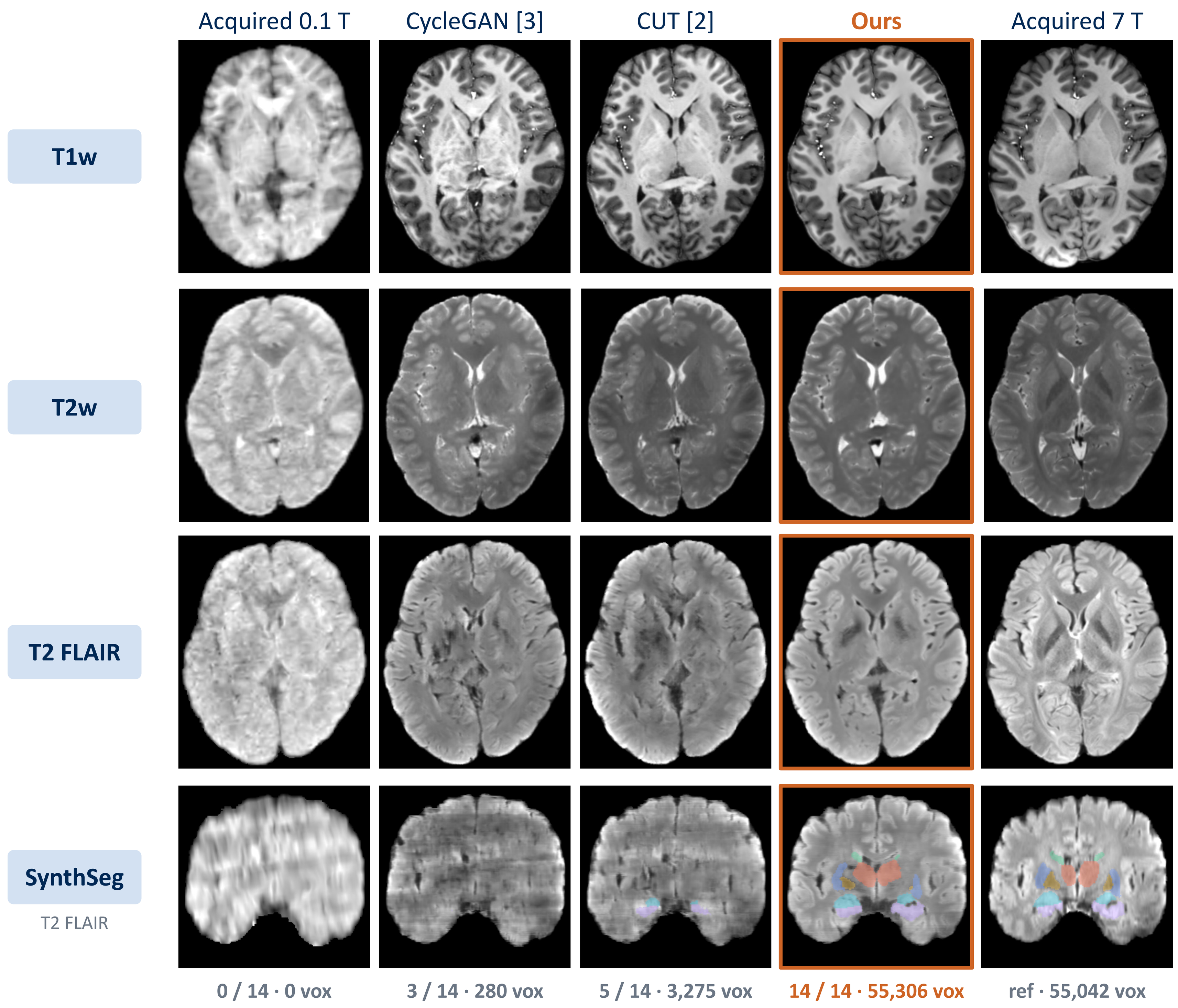
3.2 Ablation study: contextual inputs



−7.4 % nRMSE — through-plane (2.5D)
−9.1 % LPIPS — through-plane (2.5D)
+3.1 % Dice — complementary contrasts
+3.0 % VC — complementary contrasts

vs the single-channel control · all arms share seed, schedule, inference & calibration.

3.3 Qualitative analysis



Deep grey matter (thalamus, putamen): mottled and over-textured with CycleGAN & CUT — homogeneous 7 T-like with ours; on the lowest-SNR FLAIR input the baselines invent dark patches around the basal ganglia.
SynthSeg [5]: 14 / 14 deep-grey structures on our prediction, total volume within 0.5 % of the acquired 7 T — baselines mostly find the CSF-bounded hippocampus & amygdala.

4 Conclusion & discussion

CONCLUSION

- **3-step Unpaired Neural Schrödinger bridge** across 12 field × contrast cells, with a shared q_ϕ network with CUT content terms.
- **Inference-time intensity calibration** reduced nRMSE by **12.1%** and improved voxel-wise metrics.
- **Through-plane ans cross-contrast context** improved anatomical metrics: Dice **+3.1%** and volume consistency **+3.0%**.

DISCUSSION & LIMITATIONS

- **Healthy brains only** — an unpaired bridge can erase or invent lesions; clinical use needs validation on pathology.
- **Warm-starting is a compute shortcut** — every cell should also be trained at full budget to measure the gap.
- **Context enters only at fine-tuning** (first-conv inflation) — native 2.5D/3D remains unexplored.

[1] Kim et al., Unpaired image-to-image translation via Neural Schrödinger Bridge, ICLR 2024.
[2] Park et al., Contrastive learning for unpaired image-to-image translation, ECCV 2020.
[3] Zhu et al., Unpaired image-to-image translation using cycle-consistent adversarial networks, ICCV 2017.
[4] Goodfellow et al., Generative adversarial nets, NeurIPS 2014.
[5] Billot et al., SynthSeg: segmentation of brain MRI scans of any contrast and resolution, IEEE TMI 2023.
[6] Zhang et al., The unreasonable effectiveness of deep features as a perceptual metric, CVPR 2018.
[7] Wang et al., Image quality assessment: from error visibility to structural similarity, IEEE TIP 2004.

PROJECT PAGE
Code, weights & paper.