Supplementary Material: Physics-Conditioned Latent Diffusion Bridges for Ultra-Low-Field to Multi-Field MRI Synthesis

Sai Cheng[1,*], Linxin Chen[1,*], and Fangrong Zong[1,†]

[1] Beijing University of Posts and Telecommunications (BUPT), Beijing, China
[*] These authors contributed equally.
[†] Corresponding author.
chengsai@bupt.edu.cn
[2] linxin_chen@bupt.edu.cn
[3] fangrong.zong@bupt.edu.cn

This document supplements the main workshop paper. It reports implementation details that are omitted from the main text for space, and provides additional contrast-wise qualitative comparisons. Architecture dimensions, loss definitions, and the main hyperparameter table appear in the main manuscript (Sec. 2 and Table 1).

S1 DDBM Preconditioning and Sampling

The main text defines the latent bridge objective but does not specify how the denoiser is preconditioned or how bridge coefficients are estimated. These settings follow the latent DDBM formulation used in implementation.

Bridge statistics. Source and target latent variances and their covariance are estimated offline from at most 2,048 paired slices encoded by the frozen VAE. The resulting statistics define the closed-form bridge coefficients (a_t, b_t, c_t) and the time-dependent loss weight $w(t)$ used in $\mathcal{L}_{\text{DDBM}}$.

Denoiser preconditioning. The denoiser output is rescaled using the estimated endpoint variances before the squared error in $\mathcal{L}_{\text{DDBM}}$ is evaluated. Loss weights are clipped at 10^3 to limit the influence of extreme bridge times.

Static conditioning pyramid. In addition to the transformation prompt p_{tr}, the denoiser receives a fixed multi-scale pyramid built from $[z_{\text{s}}, z_{\text{coarse}}, z_{\text{coarse}} - z_{\text{s}}]$. This pyramid is fused at every encoder scale of the U-Net and is kept constant over bridge time t.

Schedule and inference. Training uses a variance-preserving schedule with endpoint offset 10^{-4}, dropout 0.1, and bridge times sampled in $[t_{\min}, T - \epsilon]$. Reverse diffusion starts from z_{s} and applies 20 deterministic Heun steps on a Karras time grid with $\rho = 7$. Exponential moving-average weights with decay 0.999 are used at inference.

S2 Phase-wise Optimization Details

Table 1 in the main manuscript lists phase lengths, learning rates, and global loss weights. The following settings are not repeated there.

Optimizer schedule. All phases use AdamW with gradient clipping at 1.0. Public pretraining applies a 1,000-step linear learning-rate warmup followed by cosine decay to 10^{-6}; Task-2 adaptation uses a 100-step warmup under the same decay rule.

Phase-3 anchor decomposition. The paired anchor $\mathcal{L}_{\text{anchor}}$ combines image L1, SSIM, edge-SSIM, and LPIPS with component weights $(1, 0.5, 0.1, 0.1)$. Its global multiplier is linearly warmed up during public Phase 3 before Task-2 fine-tuning.

Target self-reconstruction. $\mathcal{L}_{\text{target-rec}}$ decodes the frozen target posterior mean $\mu(x_{\text{t}})$ and penalizes deviation from x_{t}. This term reduces decoder drift toward latents produced only by the bridge during joint training.

Data augmentation. Task-2 adaptation applies independent horizontal and vertical flips to paired source–target slices. No through-plane augmentation is used.

S3 Additional Qualitative Results

The main manuscript (Fig. 2) reports axial T1w synthesis for the four target field strengths on a representative validation case. This section provides additional qualitative panels for T1w, T2w, and FLAIR in sagittal and coronal views, together with axial T2w and FLAIR comparisons under the same layout: 0.1 T source at left, methods by column, and target fields (1.5, 3, 5, and 7 T) by row. Validation targets are not released, so these panels contain source and synthesized images only.

T1w. Figures S1 and S2 show sagittal and coronal T1w synthesis on the same validation case as main-paper Fig. 2.

T2w. Figures S3–S5 show axial, sagittal, and coronal T2w synthesis on the same validation case.

FLAIR. Figures S6–S8 show axial, sagittal, and coronal FLAIR synthesis on the same validation case.

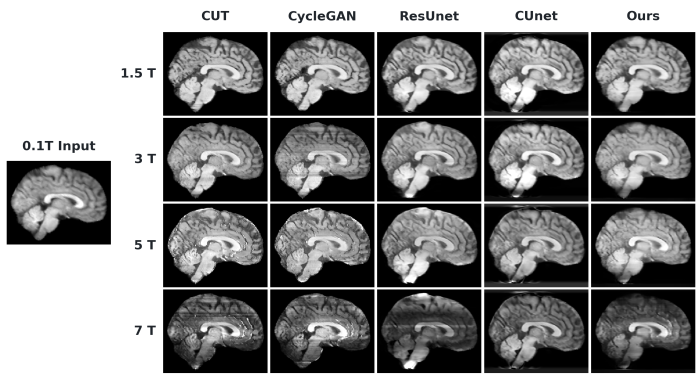

Fig. S1. Sagittal T1w synthesis on the same representative challenge validation case as main-paper Fig. 2. The leftmost image is the 0.1 T input; columns are methods and rows are target fields, including 7 T.

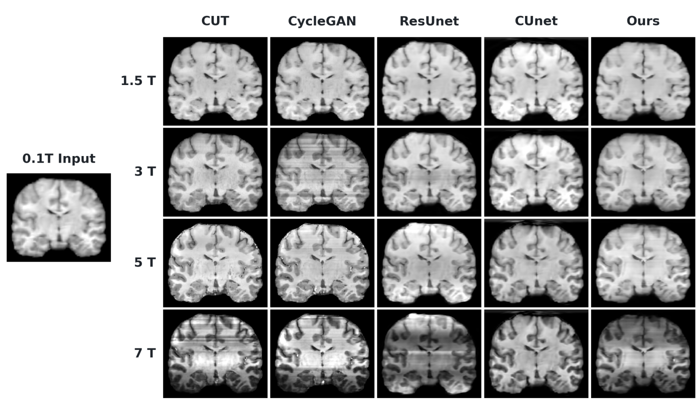

Fig. S2. Coronal T1w synthesis on the same representative challenge validation case as main-paper Fig. 2. The leftmost image is the 0.1 T input; columns are methods and rows are target fields, including 7 T.

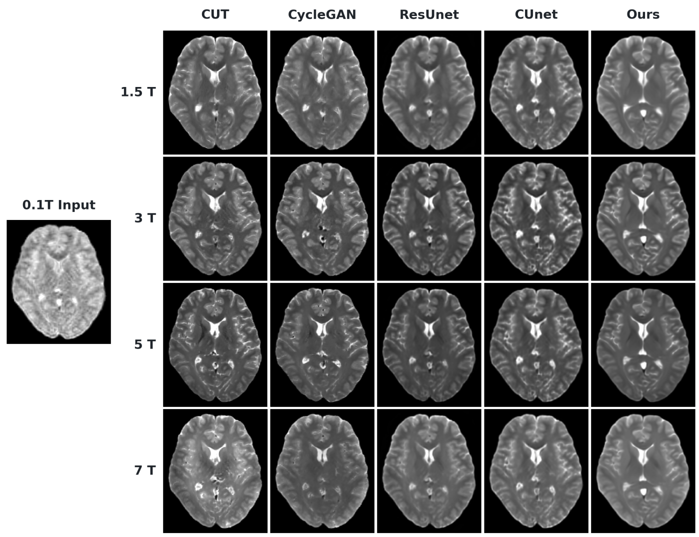

Fig. S3. Axial T2w synthesis on the same representative challenge validation case as main-paper Fig. 2. The leftmost image is the 0.1 T input; columns are methods and rows are target fields, including 7 T.

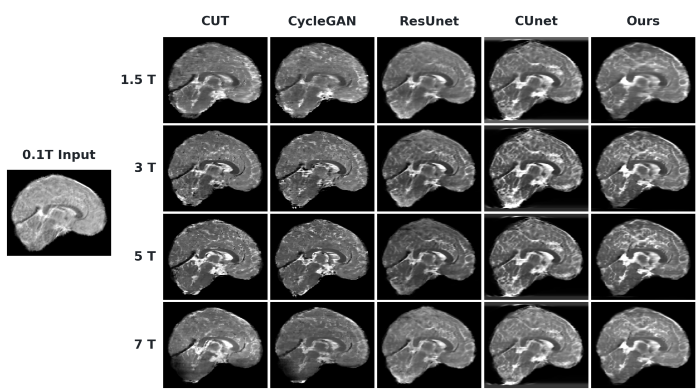

Fig. S4. Sagittal T2w synthesis on the same representative challenge validation case as main-paper Fig. 2. The leftmost image is the 0.1 T input; columns are methods and rows are target fields, including 7 T.

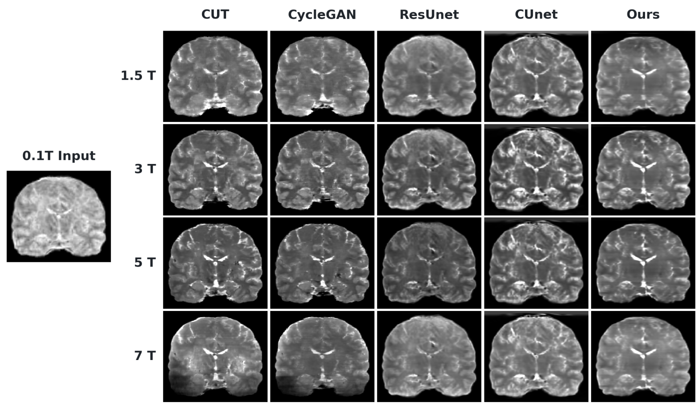

Fig. S5. Coronal T2w synthesis on the same representative challenge validation case as main-paper Fig. 2. The leftmost image is the 0.1 T input; columns are methods and rows are target fields, including 7 T.

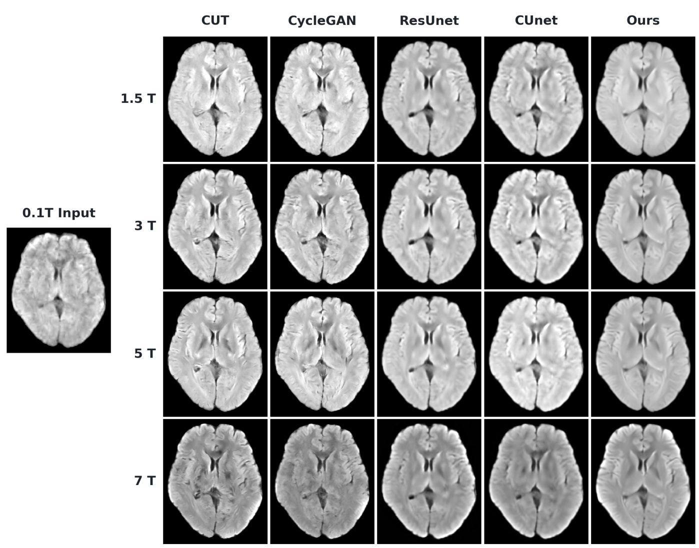

Fig. S6. Axial FLAIR synthesis on the same representative challenge validation case as main-paper Fig. 2. The leftmost image is the 0.1 T input; columns are methods and rows are target fields, including 7 T.

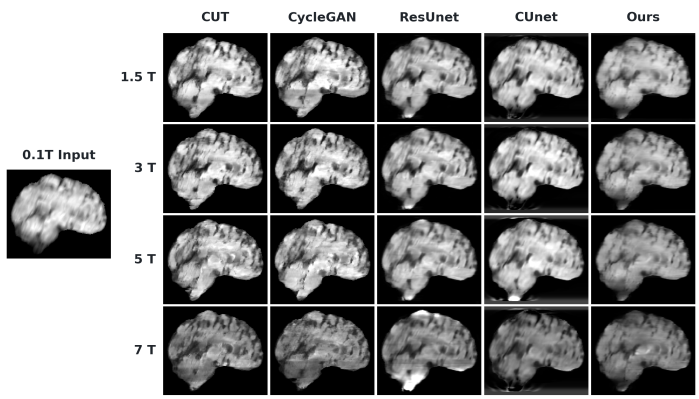

Fig. S7. Sagittal FLAIR synthesis on the same representative challenge validation case as main-paper Fig. 2. The leftmost image is the 0.1 T input; columns are methods and rows are target fields, including 7 T.

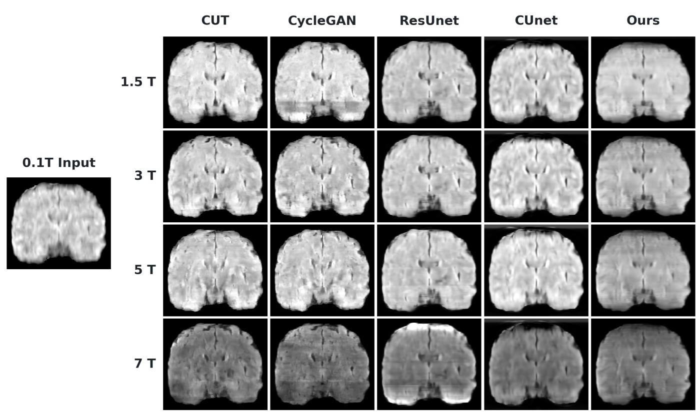

Fig. S8. Coronal FLAIR synthesis on the same representative challenge validation case as main-paper Fig. 2. The leftmost image is the 0.1 T input; columns are methods and rows are target fields, including 7 T.