This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.



## A Additional Results for reconstruction-editability trade-off experiment

Two examples illustrating the reconstruction-editability trade-off when the source image is conditioned. Increasing $\alpha$ improves source preservation but eventually heavily biases the translated image toward the source.

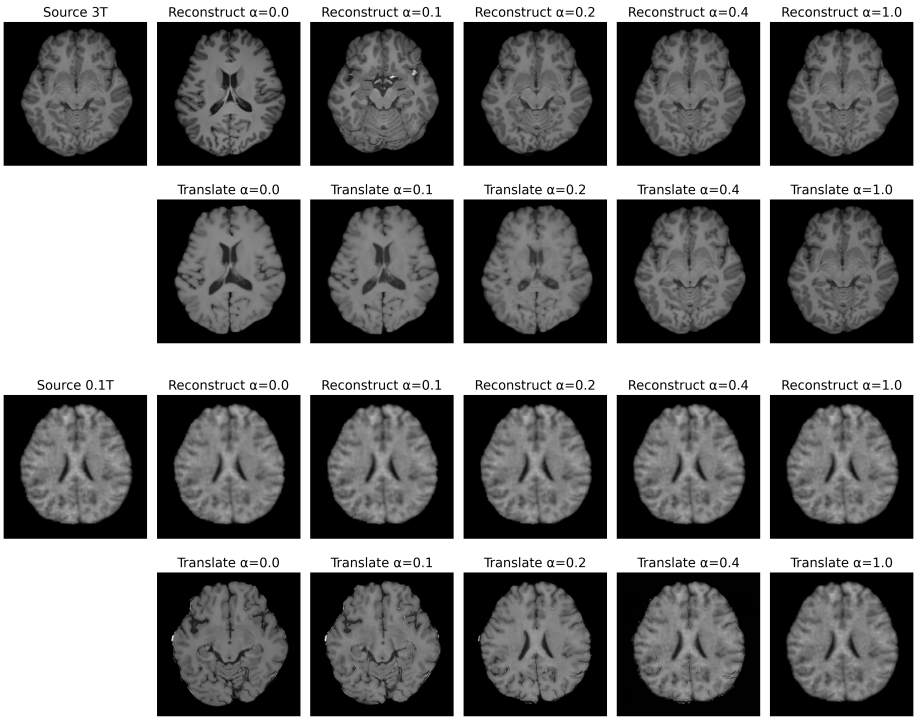


**Fig. 4.** Effect of $\alpha$-scale on DDIM reconstruction and translation (3T→1.5T, top; 0.1T→1.5T, bottom).

## B DDIM sampling time analysis

In Table 3, across all DDIM configurations, $\alpha = 0.2$ consistently improves translation performance over $\alpha = 0.0$. The (200,100) configuration achieves the best overall SSIM and LPIPS, while (50,25) reduces inference time to $1.97 \pm 0.03$ seconds per slice, resulting in a $6.81\times$ speed-up over full (200,0) inference with only a modest degradation in these perceptual metrics. We note that the nRMSE metric shows that the full trajectories (200,0) and (50,0) achieve substantially lower nRMSE (approx. 0.41 at $\alpha = 0.2$) than the corresponding truncated trajectories (approx. 0.52), suggesting that the accelerated settings preserve structural and

**Table 3.** Effect of DDIM schedule and ControlNet conditioning scale $\alpha$ on translation performance and inference time. Wall time is reported per 2D slice (batch size 1).

| DDIM | $\alpha$ | SSIM ↑ | LPIPS ↓ | nRMSE ↓ | Time (s/slice) | Speed-up |
|---|---|---|---|---|---|---|
| (200, 0) | 0.0 | $0.77 \pm 0.09$ | $0.23 \pm 0.11$ | $0.49 \pm 0.37$ | $13.4 \pm 0.08$ | 1.00× |
| | 0.2 | $0.83 \pm 0.08$ | $0.17 \pm 0.08$ | $0.41 \pm 0.36$ | | |
| (200, 100) | 0.0 | $0.82 \pm 0.08$ | $0.17 \pm 0.07$ | $0.53 \pm 0.51$ | $6.85 \pm 0.03$ | 1.96× |
| | 0.2 | $\mathbf{0.83 \pm 0.08}$ | $\mathbf{0.16 \pm 0.07}$ | $0.52 \pm 0.51$ | | |
| (50, 0) | 0.0 | $0.76 \pm 0.09$ | $0.25 \pm 0.12$ | $0.49 \pm 0.37$ | $3.64 \pm 0.06$ | 3.68× |
| | 0.2 | $0.83 \pm 0.08$ | $0.19 \pm 0.093$ | $\mathbf{0.41 \pm 0.35}$ | | |
| (50, 25) | 0.0 | $0.81 \pm 0.08$ | $0.21 \pm 0.08$ | $0.53 \pm 0.51$ | $\mathbf{1.97 \pm 0.03}$ | **6.81**× |
| | 0.2 | $0.83 \pm 0.08$ | $0.19 \pm 0.07$ | $0.52 \pm 0.51$ | | |

perceptual similarity while incurring greater pixel-wise intensity error. We use the (200,100) configuration in our study as a practical balance between translation quality and computational cost, with the intention to retain the high SSIM and LPIPS performance while approximately halving the inference time compared to full (200,0) translation, despite its higher nRMSE. Wall-clock inference time was measured on a single NVIDIA A100 GPU with batch size 1 and is reported as mean ± std over three runs per configuration.

## C Qualitative Analysis of guidance strategy and ControlNet conditioning scale $\alpha$ on field-strength translation.

The signed difference map visualises the pixel-wise change introduced by translation, computed as the difference between the source image and the translated image. Values near zero indicate little change, while positive and negative values show regions where translation, respectively, decreases or increases image intensity relative to the source.

As $\alpha$ increases, the difference maps progressively approach zero, with $\alpha$=0.5 producing outputs that remain very close to the source image.

## D Additional results for $\alpha$=0.35 and $w_{CFG} = 2.0$

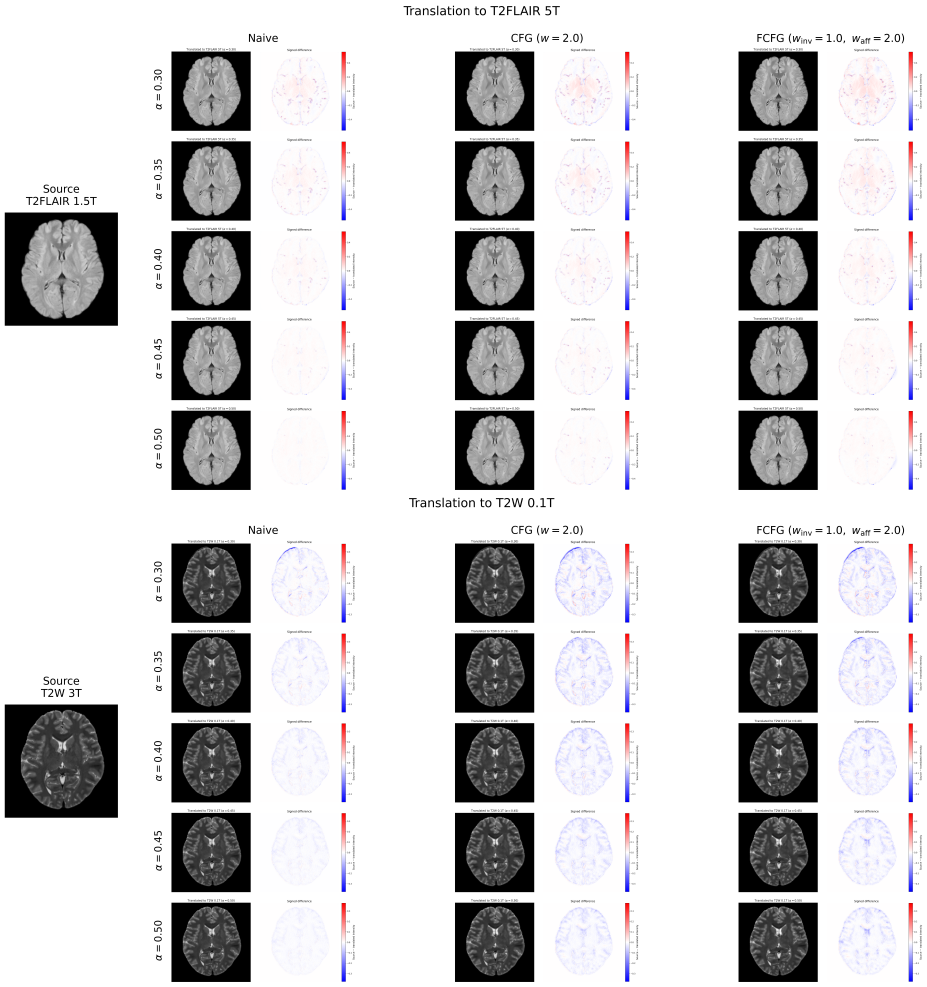


**Fig. 5.** Effect of guidance strategy and ControlNet conditioning scale $\alpha$ on field-strength translation. Translated images and signed source-to-translation difference maps are shown across $\alpha \in 0.30, 0.35, 0.40, 0.45, 0.50$ for naive conditioning, CFG ($w$=2.0), and FCFG ($w_{inv}$=1.0, $w_{aff}$=2.0). Stronger guidance produces more pronounced target-field modifications, while increasing $\alpha$ progressively reduces the difference from the source, indicating greater source preservation but weaker editability.

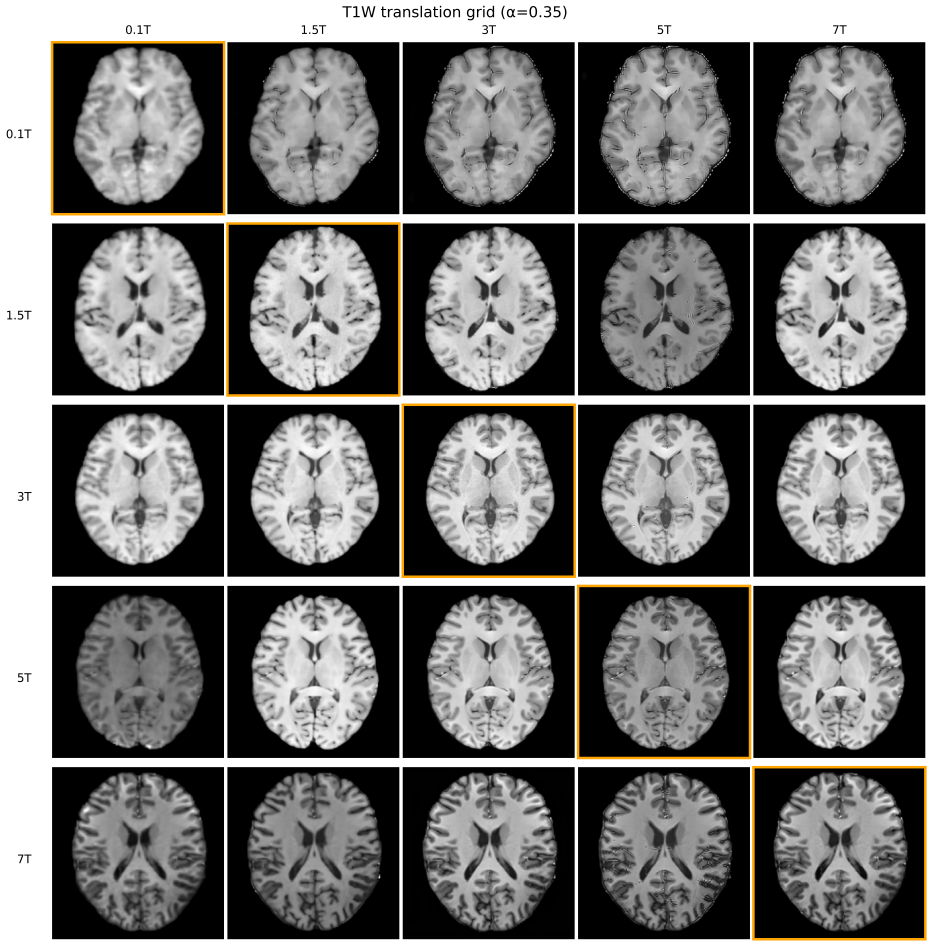


**Fig. 6.** Any-to-any field-strength translation for representative T1W samples. Yellow boxes indicate the source images. Translations are generated using ControlNet conditioning with $\alpha = 0.35$ and classifier-free guidance during conditional denoising with a guidance scale of $w = 2.0$

**Table 4.** Translation performance across source-to-target field-strength pairs ($\alpha = 0.35$, $w_{CFG} = 2.0$).

| **Pair** | **SSIM ↑** | **LPIPS ↓** | **nRMSE ↓** |
|---|---|---|---|
| 0.1T → 1.5T | $0.841 \pm 0.035$ | $0.175 \pm 0.023$ | $0.336 \pm 0.217$ |
| 0.1T → 3T | $0.811 \pm 0.043$ | $0.212 \pm 0.026$ | $0.414 \pm 0.416$ |
| 0.1T → 5T | $0.766 \pm 0.062$ | $0.239 \pm 0.032$ | $0.813 \pm 0.720$ |
| 0.1T → 7T | $0.768 \pm 0.054$ | $0.215 \pm 0.044$ | $1.155 \pm 0.619$ |
| 1.5T → 0.1T | $0.858 \pm 0.039$ | $0.145 \pm 0.028$ | $0.318 \pm 0.104$ |
| 1.5T → 3T | $0.866 \pm 0.066$ | $0.120 \pm 0.040$ | $0.404 \pm 0.548$ |
| 1.5T → 5T | $0.857 \pm 0.073$ | $0.140 \pm 0.039$ | $0.659 \pm 0.783$ |
| 1.5T → 7T | $0.852 \pm 0.080$ | $0.140 \pm 0.048$ | $0.821 \pm 0.896$ |
| 3T → 0.1T | $0.861 \pm 0.052$ | $0.136 \pm 0.036$ | $0.306 \pm 0.183$ |
| 3T → 1.5T | $0.922 \pm 0.023$ | $0.080 \pm 0.018$ | $0.202 \pm 0.101$ |
| 3T → 5T | $0.908 \pm 0.044$ | $0.089 \pm 0.024$ | $0.454 \pm 0.400$ |
| 3T → 7T | $0.889 \pm 0.062$ | $0.094 \pm 0.030$ | $0.681 \pm 0.595$ |
| 5T → 0.1T | $0.800 \pm 0.066$ | $0.178 \pm 0.032$ | $0.391 \pm 0.232$ |
| 5T → 1.5T | $0.873 \pm 0.069$ | $0.116 \pm 0.039$ | $0.290 \pm 0.214$ |
| 5T → 3T | $0.869 \pm 0.098$ | $0.113 \pm 0.054$ | $0.357 \pm 0.272$ |
| 5T → 7T | $0.896 \pm 0.043$ | $0.089 \pm 0.025$ | $0.519 \pm 0.359$ |
| 7T → 0.1T | $0.767 \pm 0.059$ | $0.202 \pm 0.051$ | $0.494 \pm 0.145$ |
| 7T → 1.5T | $0.824 \pm 0.088$ | $0.165 \pm 0.064$ | $0.400 \pm 0.208$ |
| 7T → 3T | $0.827 \pm 0.083$ | $0.168 \pm 0.062$ | $0.495 \pm 0.152$ |
| 7T → 5T | $0.814 \pm 0.118$ | $0.162 \pm 0.066$ | $0.712 \pm 0.880$ |