Autoregressive Image Diffusion: Generation of Image Sequences and Application in MRI Reconstruction

Guanxiong Luo[1] Shoujin Huang Martin Uecker[1,2]

[1]Institute for Diagnostic and Interventional Radiology, University Medical Center Göttingen
[2]Institute of Biomedical Imaging, Graz University of Technology, Graz, Austria

Introduction

In this work, we present the autoregressive image diffusion (AID) model for image sequences and use it to sample the posterior for accelerated MRI reconstruction. The algorithm incorporates both undersampled k-space and pre-existing information. The results show that the AID model can robustly generate sequentially coherent image sequences. In MRI applications, the AID can outperform the standard diffusion model and reduce hallucinations, due to the learned inter-image dependencies. The project code is available at `https://github.com/mrirecon/aid`.

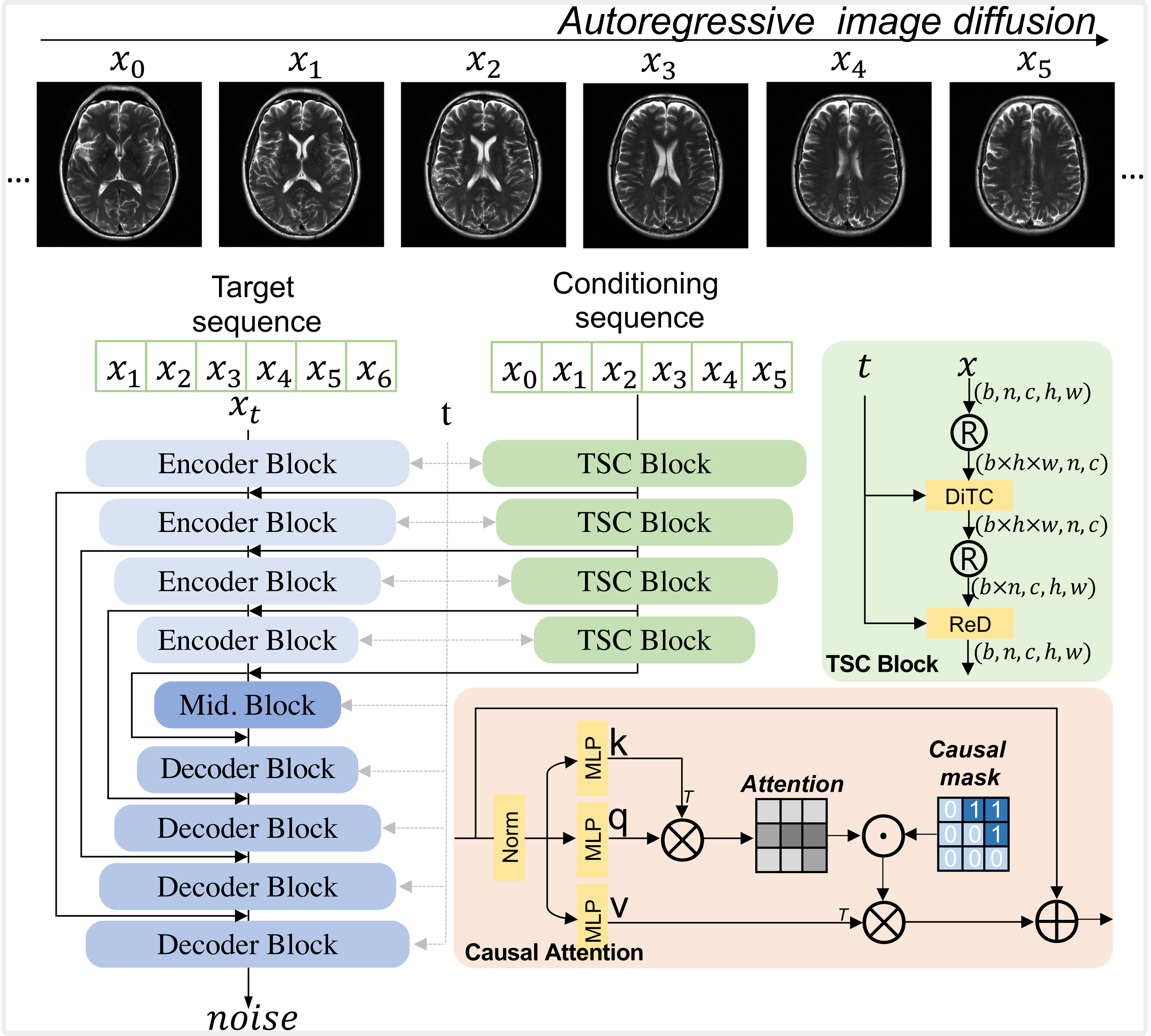

Figure 1. The interaction between the images in conditioning sequence occurs in the DiT-Block, which has a causal attention module to ensure x_n is conditioned on previous images $x_{<n}$. During training, the net predicts the noise for each noisy image that is sampled from the target sequence given the conditioning sequence in parallel. During generation, the net iteratively refines the noisy input to produce a clean image, which is then appended to the conditioning sequence.

Theory

Given a sequence of k-space $\mathbf{y} = \{y_1, \ldots, y_N\}$, each posterior in $\{p_\theta(x_n|y_n, x^0_{<n}) | 1 < n < N\}$ is expressed as

$$p_\theta(x_n|y_n, x^0_{<n}) = \frac{p(y_n|x_n, x^0_{<n})\, p_\theta(x_n|x^0_{<n})}{p(y_n|x^0_{<n})} = \frac{p(y_n|x_n)\, p_\theta(x_n|x^0_{<n})}{p(y_n)} \propto p(y_n|x_n)\, p_\theta(x_n|x^0_{<n}) \,, \quad (1)$$

when the acquisition of y_n is independent of the image $x^0_{<n}$, y_n and $x^0_{<n}$ are conditionally independent given x_n. Then, we have

$$p_\theta(x_n^{t-1}|x_n^t, y_n, x^0_{<n}) \propto p(y_n|x_n^t)\, p_\theta(x_n^{t-1}|x_n^t, x^0_{<n}) \,. \quad (2)$$

Samples

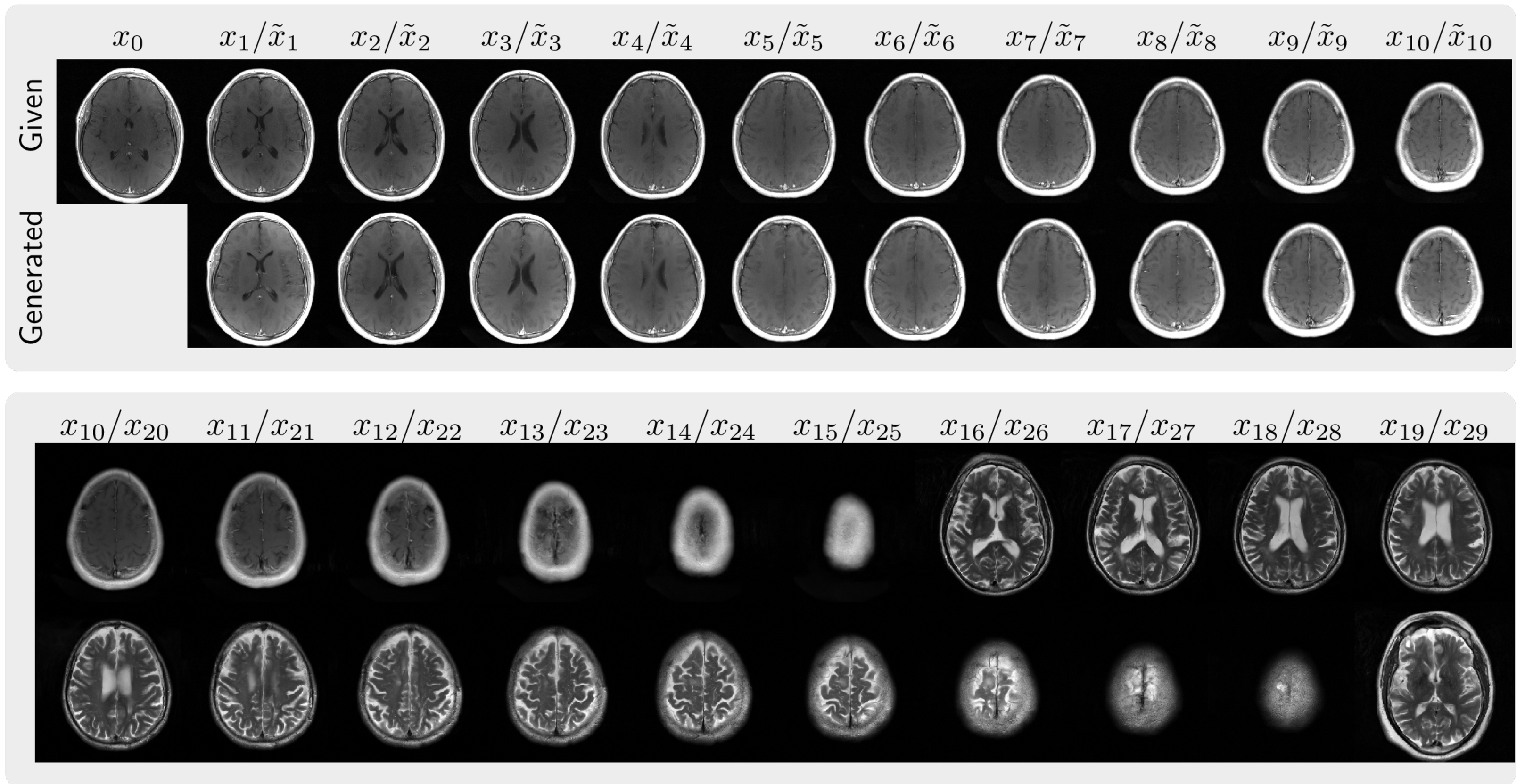

Figure 2. (a): A sequence of images from dataset is shown in the first row and is used as conditioning to generate retrospective samples that are shown in the second row. (b): With the given sequence in (a) as a warm start, prospective samples extending it are shown.

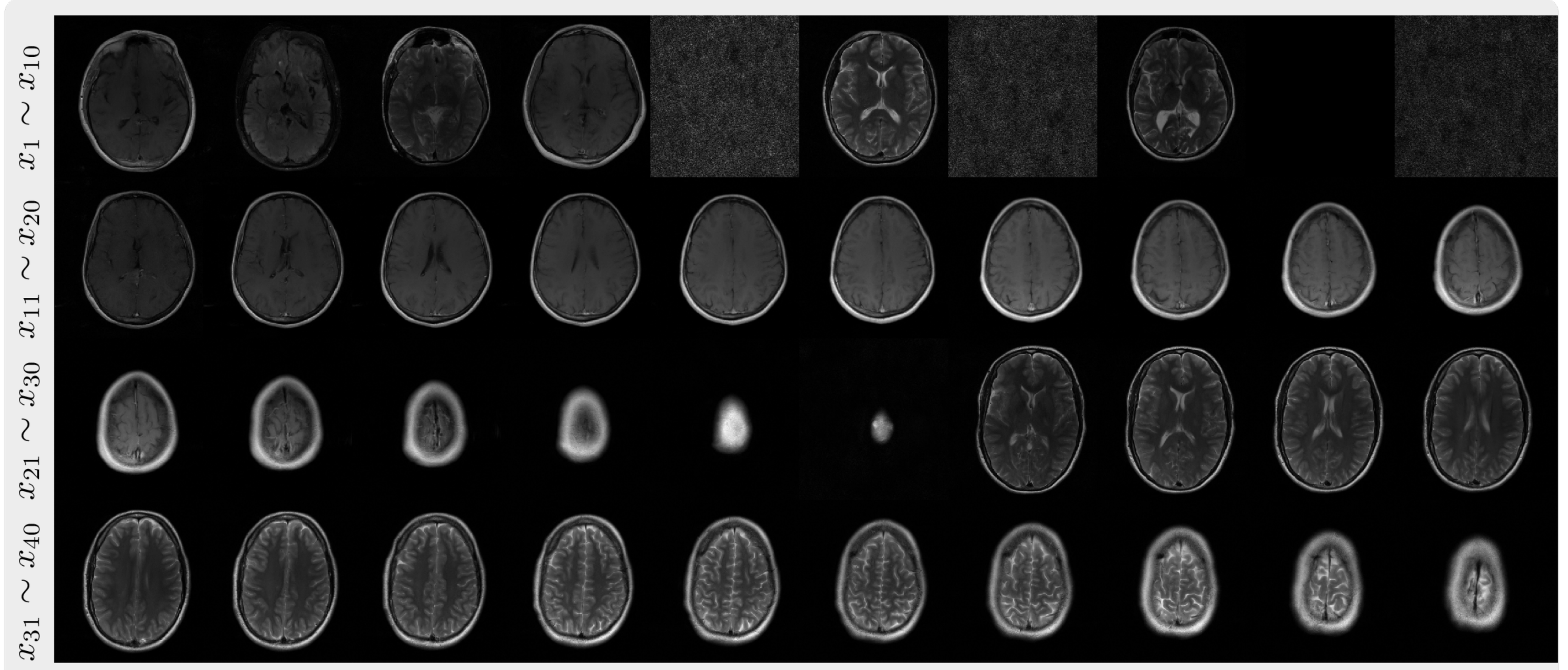

Figure 3. Prospective samples with cold start. The initial images generated in the cold start are not sequentially coherent, but as the sampling process continues, the model progressively generates more sequentially coherent and realistic images.

MRI Reconstruction

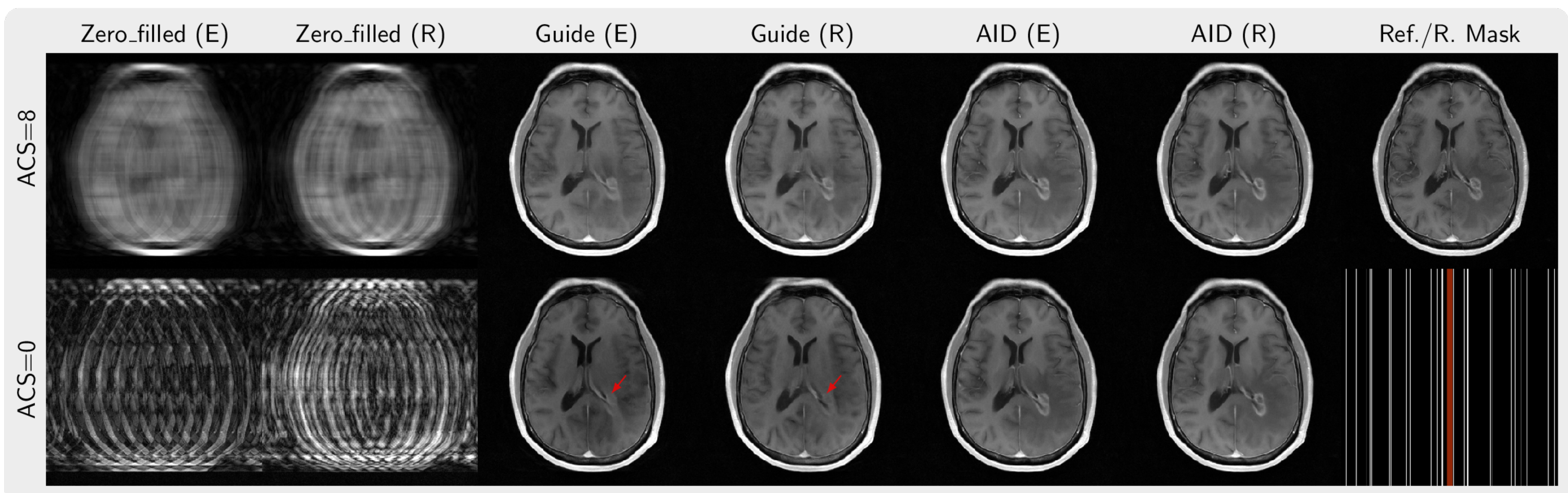

Figure 4. E: equispaced, R: random. The last column shows the reference and the random sampling mask in k-space. The red lines are autocalibration signal (ACS) and equispaced mask is not shown. Zero-filled images are computed by inverse Fourier transform of the zero-filled k-space data. The hallucinations are pointed with red arrows.

Unfolding the Posterior

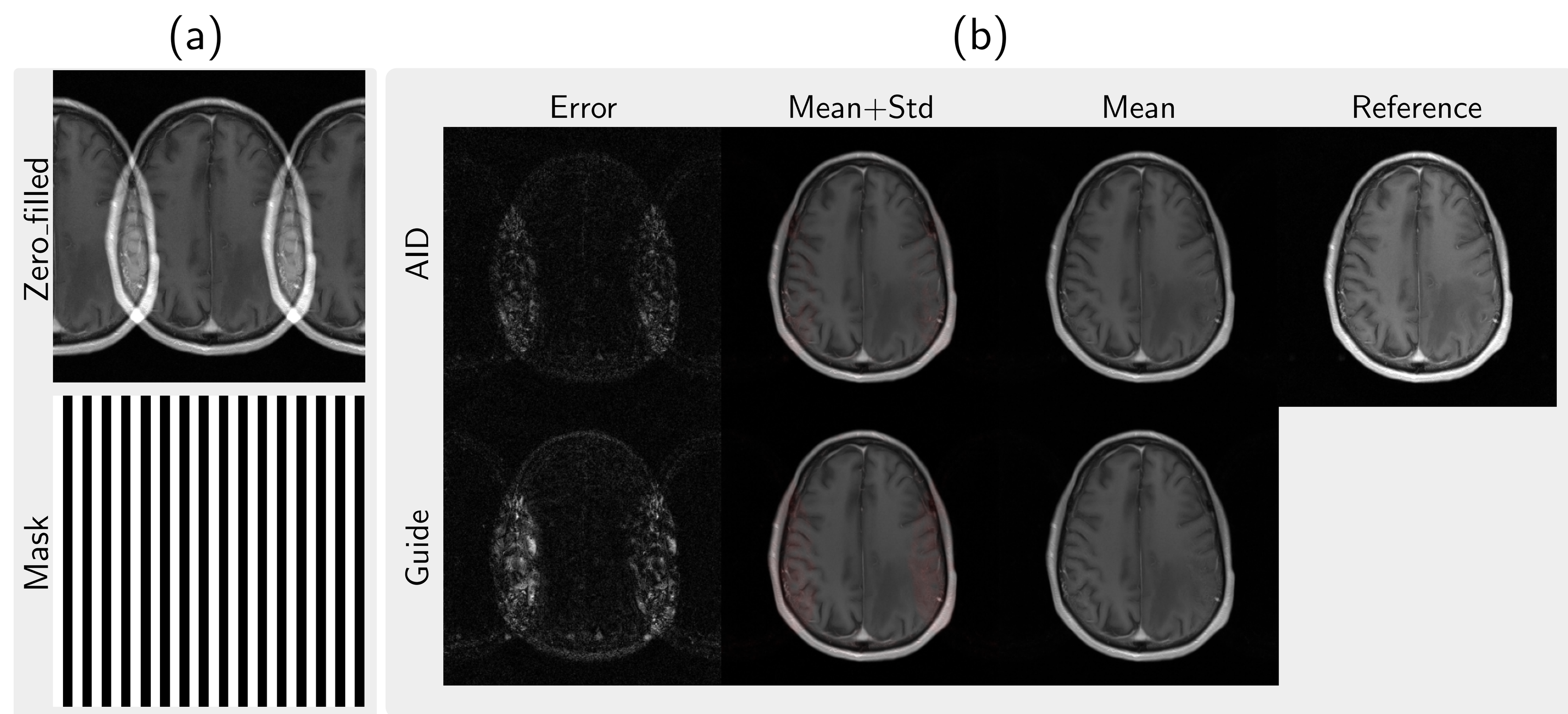

Figure 5. (a): The folded single-coil image caused by two-times undersampling mask. (b): The comparison of unfolding ability by the autoregressive and the standard diffusion model, i.e., AID (top) and Guide (bottom). Reference image is reconstructed from k-space data without undersampling. The error is the difference between the mean, x_{MMSE}, and the reference image. The "Mean+std" is the mean highlighted with confidence interval, which indicates the reconstruction by AID is more trustworthy in the region of folding artifacts.

Guided vs AID

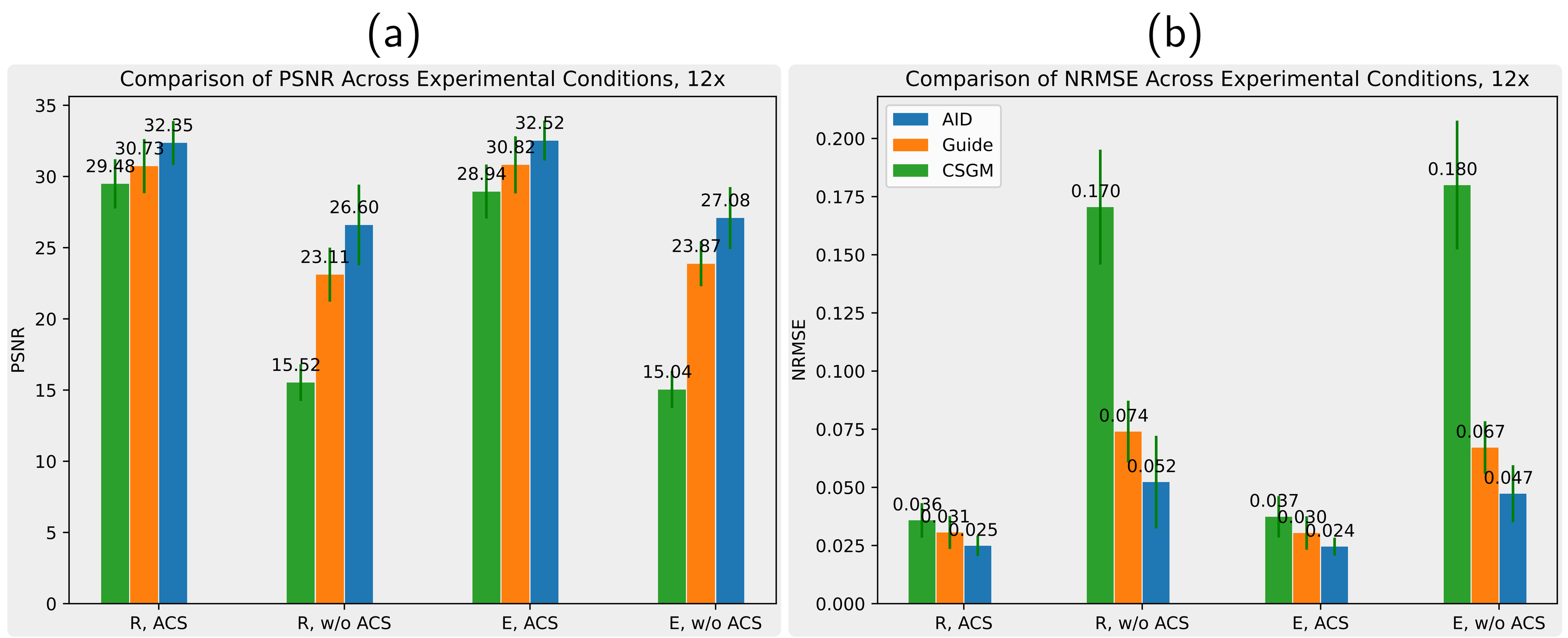

Figure 6. E: equispaced, R: random. (a): PSNR and (b): NRMSE of the images reconstructed from the twelve-times undersampled k-space data using the autoregressive diffusion model (AID), the standard diffusion model (Guide), and the baseline method CSGM. PSNR higher is better, and NRMSE lower is better.

Conclusion

The proposed autoregressive image diffusion model offers an approach to generating image sequences, with significant potential as a trustworthy prior in accelerated MRI reconstruction. In various experiments, it outperforms the standard diffusion model in terms of both image quality and robustness by taking the advantage of the prior information on inter-image dependencies.

Reference

1. Song et al. (2019, NeurIPS) 2. Song et al. (2021, ICLR) 3. Jalal et al. (2021, NeurIPS) 4. Peebles et al. (2023, CVPR) 5. Luo et al. (2023, MRM) 6. Chung et al. (2023, MIA) 7. Luo et al. (2024, NeurIPS)

 luoguan5@gmail.com