Supplementary Materials for: Unsupervised MRI Reconstruction via Zero-Shot Learned Adversarial Transformers

Yilmaz Korkmaz[1,2], Salman UH Dar[1,2], Mahmut Yurt[1,2], Muzaffer Özbey[1,2], and Tolga Çukur[1,2,3]

[1] Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey
[2] National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara 06800, Turkey
[3] Neuroscience Program, Bilkent University, Ankara 06800, Turkey

Contents

I. Supplementary Text

A. Positional encoding variables

For input feature maps $X \in \mathbb{R}^{h_1 \times h_2 \times u}$, sinusoidal position encoding variables $PE \in \mathbb{R}^{h_1 \times h_2 \times u}$ are set at location (loc_{h_1}, loc_{h_2}, loc_u) as [74]:

$$PE[loc_{h_1}, loc_{h_2}, loc_u] = \begin{cases} sin\left(\frac{loc_{h_2}}{10000^{4(\frac{loc_u}{u})}}\right) & 0 < loc_u \leq \frac{u}{4} \\ cos\left(\frac{loc_{h_2}}{10000^{4(\frac{loc_u}{u} - \frac{1}{4})}}\right) & \frac{u}{4} \leq loc_u \leq \frac{u}{2} \\ sin\left(\frac{loc_{h_1}}{10000^{4(\frac{loc_u}{u} - \frac{2}{4})}}\right) & \frac{u}{2} \leq loc_u \leq \frac{u}{3} \\ cos\left(\frac{loc_{h_1}}{10000^{4(\frac{loc_u}{u} - \frac{3}{4})}}\right) & \frac{u}{3} \leq loc_u \leq u \end{cases}$$

where loc_{h_1} and loc_{h_2} lie in range [-1, 1], covering complete field of view along the first two spatial dimensions, and loc_u is the channel index.

B. Architectural Details

1. Synthesizer:

- Layer 1 (4x4): Input(Constant) → Cross-Attention Transformer Block → Output
- Layer 2 (8x8): Input → Upsample→ Cross-Attention Transformer Block + Upsample(Input) → Output
- Layer 3 (16x16): Input → Upsample→ Cross-Attention Transformer Block + Upsample(Input) → Output
- Layer 4 (32x32): Input → Upsample → Cross-Attention Transformer Block + Upsample(Input) → Output
- Layer 5 (64x64): Input → Upsample→ Cross-Attention Transformer Block + Upsample(Input) → Output
- Layer 6 (128x128): Input → Upsample→ Cross-Attention Transformer Block + Upsample(Input) → Output
- Layer 7 (256x256): Input → Upsample→ Modulated Convolution → Output
- Cross-attention Transformer Block: Input → Cross-Attention + Noise → Modulated Convolution → Cross-Attention + Noise → Output

2. Mapper:

a) ***Local Stream***:

- Layer 1: Input → Self-Attention Block → Output
- Layer 2: Input → Self-Attention Block → Output
- Layer 3: Input → Self-Attention Block → Output
- Layer 4: Input → Self-Attention Block → Output
- Layer 5: Input → Fully-connected→ Output
- Self-Attention Block: Input → Self-Attention → Fully-connected → Fully-connected + Input →Output

b) ***Global Stream***:

- Layer 1: Input → Fully-connected → Output
- Layer 2: Input → Fully-connected → Output
- Layer 3: Input → Fully-connected → Output
- Layer 4: Input → Fully-connected → Output
- Layer 5: Input → Fully-connected → Output
- Layer 6: Input → Fully-connected → Output
- Layer 7: Input → Fully-connected → Output
- Layer 8: Input → Fully-connected → Output
- Layer 9: Input → Fully-connected → Output

3. Discriminator:

- Layer 1 (256x256): Input → Convolution → Downsample + Downsample(Input) → Output
- Layer 2 (128x128): Input → Convolution → Downsample + Downsample(Input) → Output
- Layer 3 (64x64): Input → Convolution → Downsample + Downsample(Input) → Output
- Layer 4 (32x32): Input → Convolution → Downsample + Downsample(Input) → Output
- Layer 5 (16x16): Input → Convolution → Downsample + Downsample(Input) → Output
- Layer 6 (8x8): Input → Convolution → Downsample + Downsample(Input) → Output
- Layer 7 (4x4): Input → Convolution → Downsample + Downsample(Input) → Output

II. Supplementary Tables

Supp. Table I: Within-domain reconstruction performance for T_1- and T_2-weighted acquisitions in the IXI dataset at R=4 and 8.

	LORAKS		**GAN$_{\text{sup}}$**		**SSDU**		**GAN$_{\text{prior}}$**		**SAGAN**		**SLATER**	
	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)
T_1, R=4	30.7±1.2	91.7±1.0	37.5±0.5	97.8±0.2	37.9±0.6	97.8±0.2	34.4±0.8	94.4±0.7	32.1±0.9	92.1±0.7	38.8±0.8	97.9±0.5
T_1, R=8	26.8±0.9	87.3±1.1	33.3±0.6	95.7±0.3	33.1±0.7	93.9±0.7	29.3±1.2	89.7±1.4	28.6±0.9	88.3±1.2	33.2±0.9	95.2±0.9
T_2, R=4	35.4±0.5	92.3±1.2	38.7±0.8	96.8±0.3	38.9±0.7	96.3±0.4	33.4±0.9	87.5±1.0	34.9±0.6	91.6±1.1	40.0±0.8	97.7±0.5
T_2, R=8	31.4±0.4	88.2±1.3	34.2±0.8	94.3±0.6	33.7±0.9	91.6±1.1	31.2±0.7	85.3±1.0	30.7±0.5	86.4±1.4	34.1±0.8	94.8±0.7

Supp. Table II: Across-domain reconstruction performance for T_1- and T_2-weighted acquisitions in the IXI and fastMRI datasets. In A->B, A and B denote the acceleration rates in training versus test domains. Because LORAKS is untrained, and GAN$_{\text{prior}}$, SAGAN and SLATER do not make any assumptions regarding the imaging operator during training, their across-domain reconstruction performance is equivalent to the within-domain performance for the target acceleration rate.

	LORAKS		**GAN$_{\text{sup}}$**		**SSDU**		**GAN$_{\text{prior}}$**		**SAGAN**		**SLATER**	
	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)
IXI T_1, 8->4	30.7±1.2	91.7±1.0	32.8±0.9	96.6±0.3	33.1±1.2	94.5±0.9	34.4±0.8	94.4±0.7	32.1±0.9	92.1±0.7	38.8±0.8	97.9±0.5
IXI T_2, 8->4	35.4±0.5	92.3±1.2	33.7±0.5	93.8±0.4	34.8±0.8	92.4±1.0	33.4±0.9	87.5±1.0	34.9±0.6	91.6±1.1	40.0±0.8	97.7±0.5
fastMRI T_1, 8->4	33.4±2.7	82.2±7.7	34.8±2.0	93.7±5.7	35.0±2.5	92.1±7.4	32.8±2.0	92.5±5.2	36.1±2.6	94.1±5.1	37.6±3.2	93.9±9.5
fastMRI T_2, 8->4	34.3±1.0	90.8±1.6	33.3±1.0	94.8±0.6	32.0±1.9	92.5±1.6	33.5±1.1	91.5±1.8	33.5±1.3	94.1±0.8	36.3±1.2	95.5±0.7

Supp. Table III: Reconstruction performance in ablation experiments for SLATER. Metrics are reported for T_1- and T_2-weighted acquisitions in the IXI dataset at R=4.

	None		**Latent**		**Latent+Noise**		**Latent+Noise+Weight**	
	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)
T_1	26.7±1.3	87.1±1.2	32.4±1.1	94.1±0.7	34.0±1.2	96.6±0.4	38.8±0.8	97.9±0.5
T_2	30.4±0.7	80.5±1.4	32.9±0.8	88.0±0.9	36.2±0.8	94.3±0.6	40.0±0.8	97.7±0.5

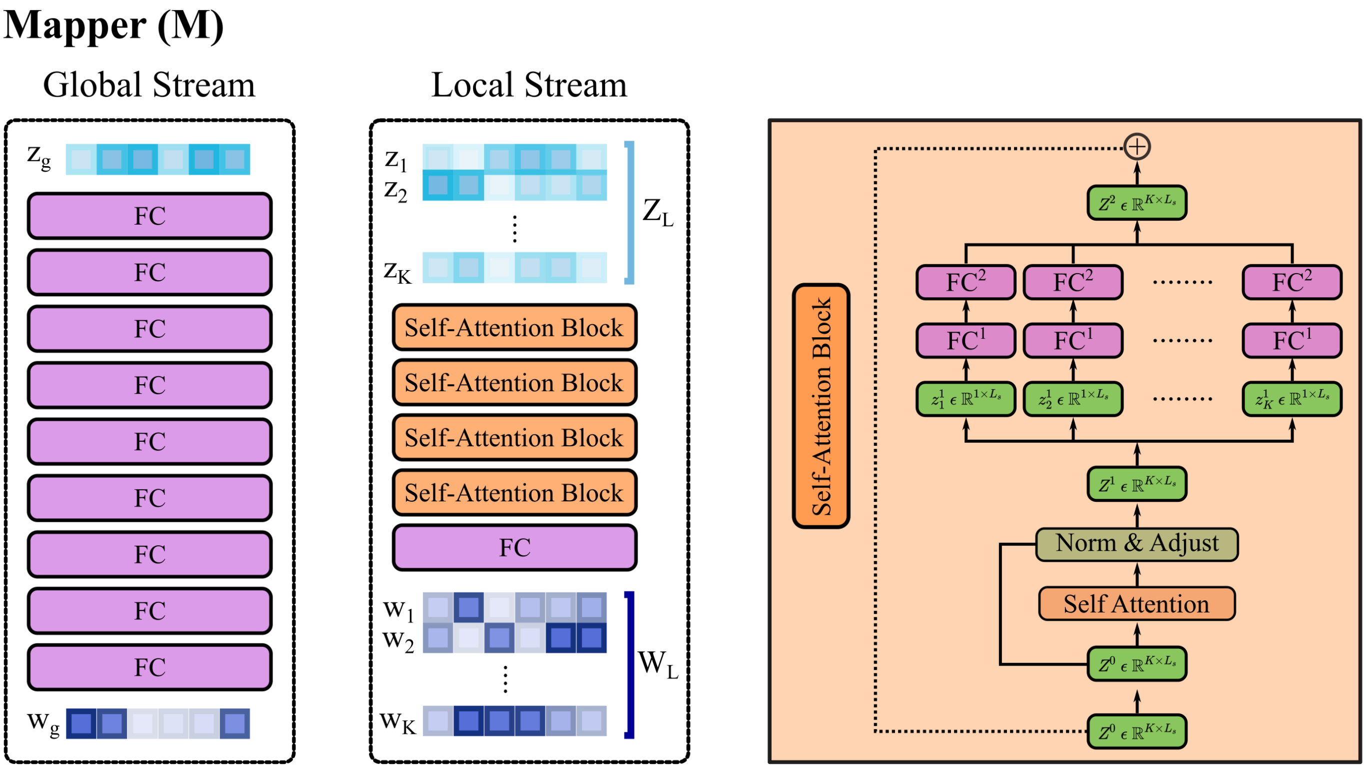
Supp. Table IV: Average training time of models in min:sec format per epoch in the IXI dataset. Note that LORAKS does not perform any training.

	LORAKS	**GAN$_{\text{sup}}$**	**SSDU**	**GAN$_{\text{prior}}$**	**SAGAN**	**SLATER**
Time (min:sec)	–	6:49	1:49	6:22	6:49	8:10

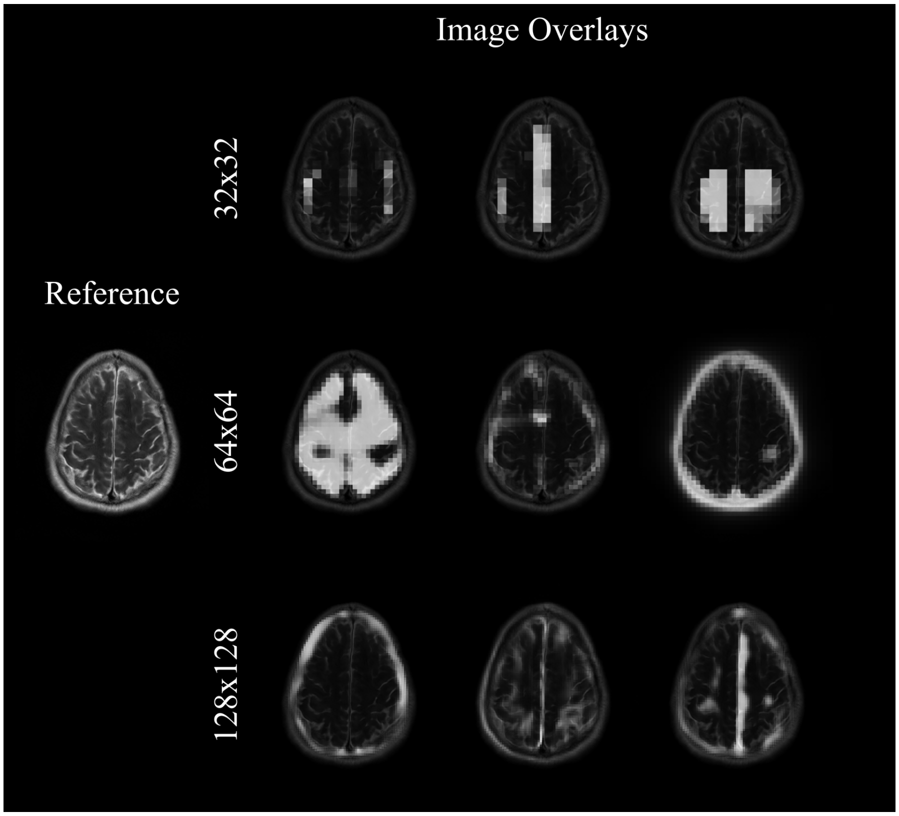
Supp. Table V: Reconstruction performance for T$_1$- and T$_2$-weighted acquisitions in the IXI dataset at R=4 and 8 based on the weight propagation procedure. Note that weight propagation only affects the performance of GAN$_{\text{prior}}$ and SLATER for which weight optimization is performed during inference.

	LORAKS		**GAN$_{\text{sup}}$**		**SSDU**		**GAN$_{\text{prior}}$**		**SAGAN**		**SLATER**	
	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)	PSNR	SSIM(%)
T1, R=4	30.7±1.2	91.7±1.0	37.5±0.5	97.8±0.2	37.9±0.6	97.8±0.2	34.15±0.93	95.06±0.55	32.1±0.9	92.1±0.7	38.63±0.88	98.17±0.24
T1, R=8	26.8±0.9	87.3±1.1	33.3±0.6	95.7±0.3	33.1±0.7	93.9±0.7	29.02±1.11	88.91±1.41	28.6±0.9	88.3±1.2	33.04±1.05	96.06±0.52
T2, R=4	35.4±0.5	92.3±1.2	38.7±0.8	96.8±0.3	38.9±0.7	96.3±0.4	33.04±0.84	88.39±1.23	34.9±0.6	91.6±1.1	39.80±0.80	97.77±0.27
T2, R=8	31.4±0.4	88.2±1.3	34.2±0.8	94.3±0.6	33.7±0.9	91.6±1.1	30.83±0.64	87.30±1.17	30.7±0.5	86.4±1.4	33.96±0.77	94.10±0.46

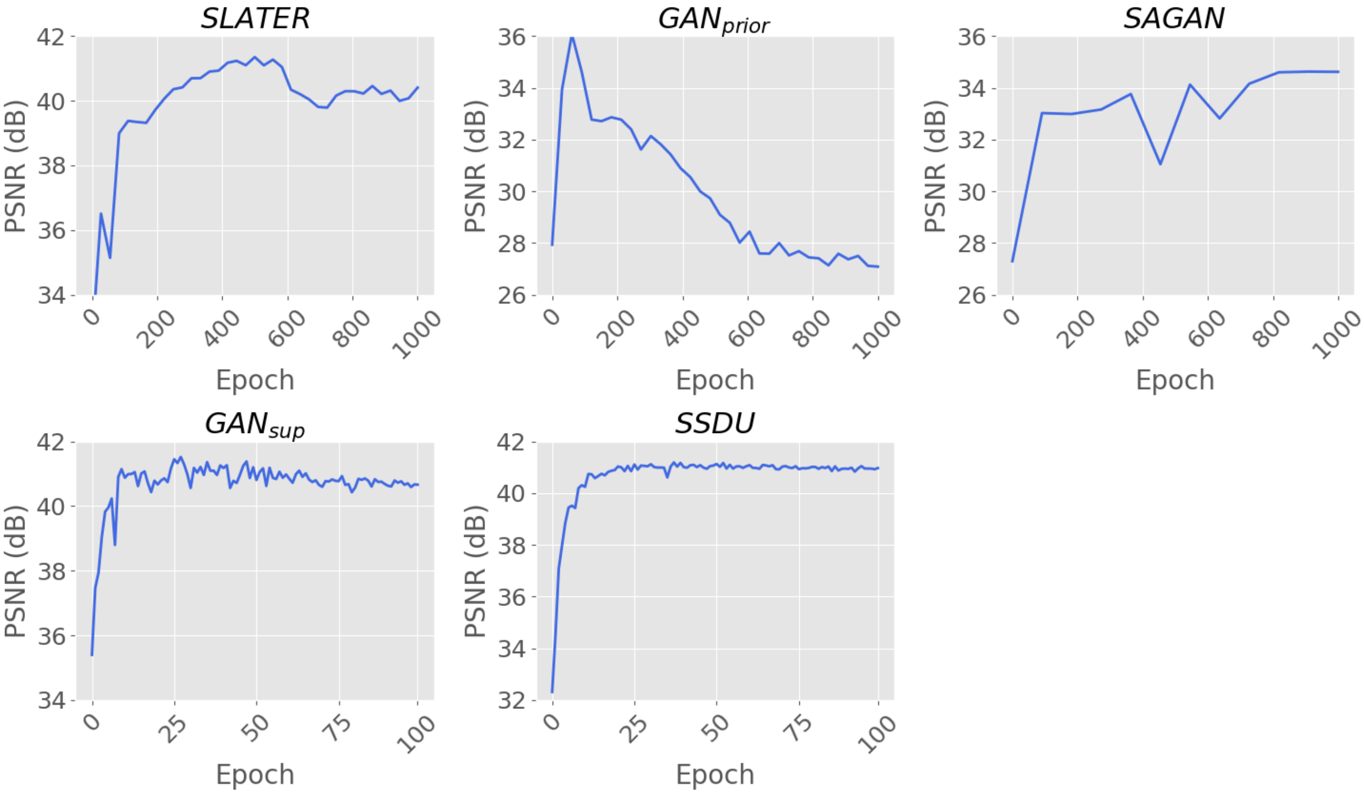
III. Supplementary Figures


Mapper (M)
Global Stream
Local Stream
z_g
FC
FC
FC
FC
FC
FC
FC
FC
FC
w_g
z_1
z_2
z_K
Z_L
Self-Attention Block
Self-Attention Block
Self-Attention Block
Self-Attention Block
FC
w_1
w_2
w_K
W_L
Self-Attention Block
$Z^2 \epsilon \mathbb{R}^{K \times L_s}$
FC^2
FC^2
FC^2
FC^1
FC^1
FC^1
$z_1^1 \epsilon \mathbb{R}^{1 \times L_s}$
$z_2^1 \epsilon \mathbb{R}^{1 \times L_s}$
$z_K^1 \epsilon \mathbb{R}^{1 \times L_s}$
$Z^1 \epsilon \mathbb{R}^{K \times L_s}$
Norm & Adjust
Self Attention
$Z^0 \epsilon \mathbb{R}^{K \times L_s}$
$Z^0 \epsilon \mathbb{R}^{K \times L_s}$

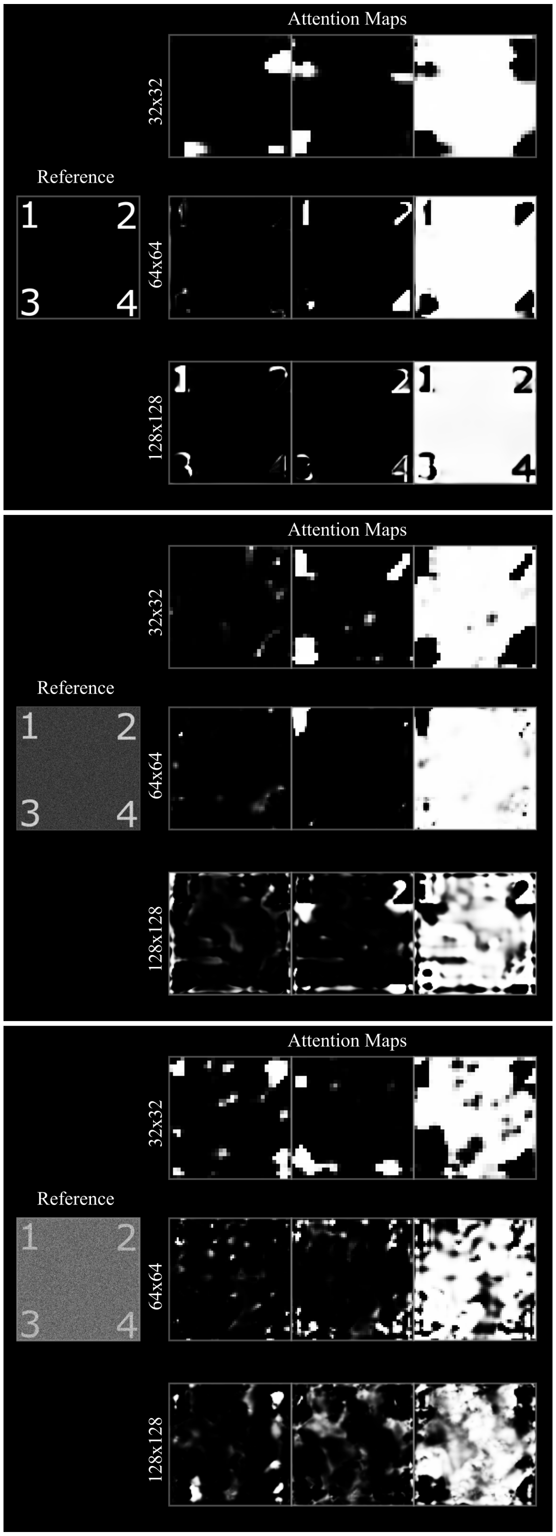
Supp. Fig. 1: Mapper is a multi-layered architecture comprising two separate processing streams: a global stream dedicated to the global latent variable w_g, and a local stream dedicated to the local latent variables $W_l = \{w_1, w_2, ..., w_K\}$. The global stream contains a cascade of fully-connected sub-blocks. Meanwhile, the local stream is a cascade of self-attention sub-blocks followed by a fully-connected sub-block (see rightmost panel for the architecture of the self-attention sub-block). Self-attention sub-blocks enable interactions among individual local latents.



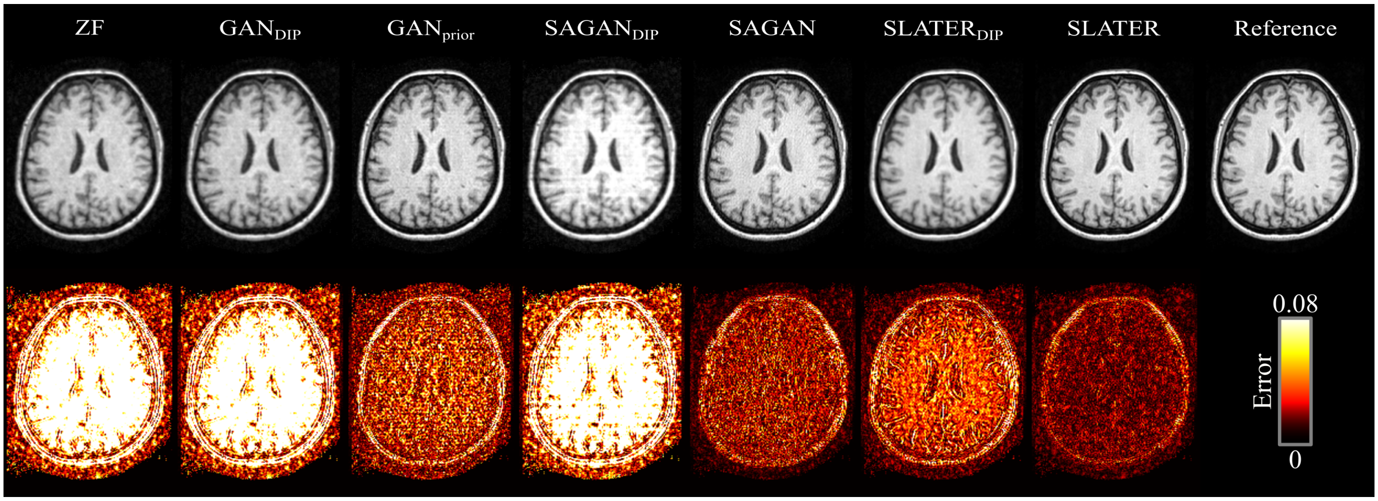
Supp. Fig. 2: Cross-attention maps in SLATER for a T_2-weighted acquisition. Sample attention maps from the first cross-attention transformer block are displayed across three resolutions (i.e., 32x32, 64x64, 128x128 at network layers 4-6). At each resolution, respective maps are displayed in overlaid format onto the MR image, and the reference MR image is also shown. Attention maps for separate latents show segregated spatial distribution. They also tend to group tissue clusters with similar signal intensity and texture, where the clusters are broadly distributed across the image and they are often spatially noncontiguous.



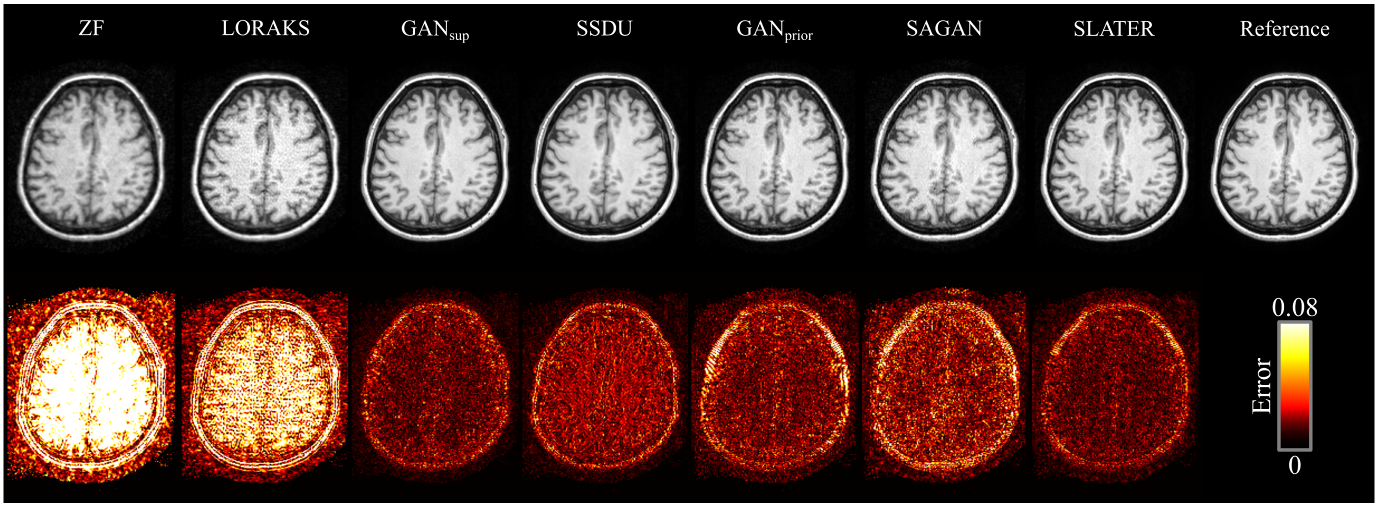
Supp. Fig. 3: Reconstruction performance in the validation set as a function of number of training epochs. Results from supervised (GAN$_{\text{sup}}$) and unsupervised models (SSDU, GAN$_{\text{prior}}$, SAGAN and SLATER) are shown for T_1-weighted acquisitions in IXI at R=4. For unsupervised models, hyperparameter selection in the validation set was actually performed based on the difference between recovered and acquired k-space samples in undersampled data. However, to facilitate interpretation, here performance for all methods is displayed as PSNR between reconstructed and ground-truth images.



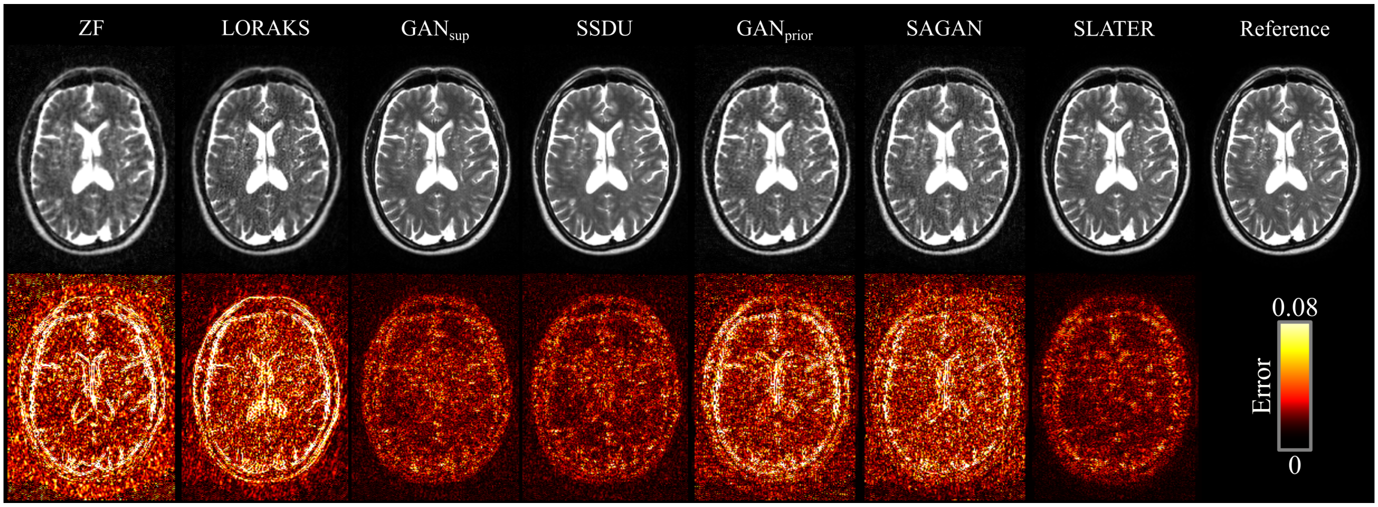
Supp. Fig. 4: Cross-attention maps in SLATER for a simulated digit phantom with varying levels of noise. Relative to a peak signal intensity of 1, top, middle and bottom panels display sample attention maps for no noise, noise variance of 0.01, and noise variance of 0.1, respectively. Within each panel, maps from the first cross-attention sub-block are shown at three resolutions (i.e., 32x32, 64x64, 128x128 at network layers 4-6), along with the reference phantom images.



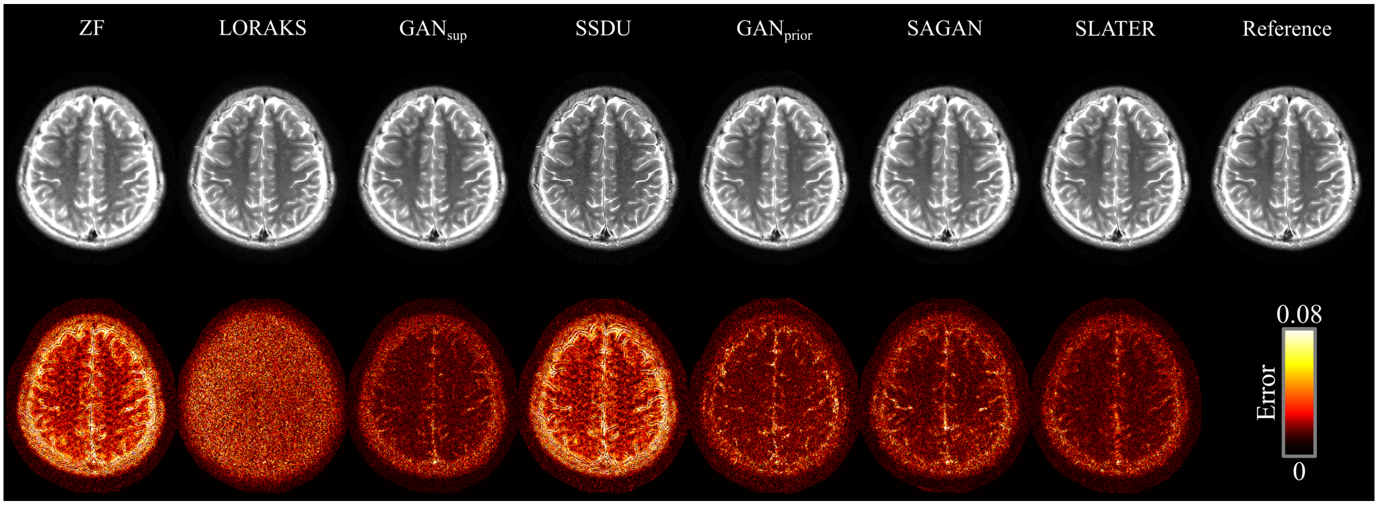
Supp. Fig. 5: Reconstructions of a representative T_1-weighted acquisition at R=4 are shown for the Fourier method (ZF), DIP methods (GAN$_{DIP}$, SAGAN$_{DIP}$, SLATER$_{DIP}$) and zero-shot reconstructions (GAN$_{prior}$, SAGAN, SLATER) along with the reference image. Corresponding error maps are underneath the images for each method.



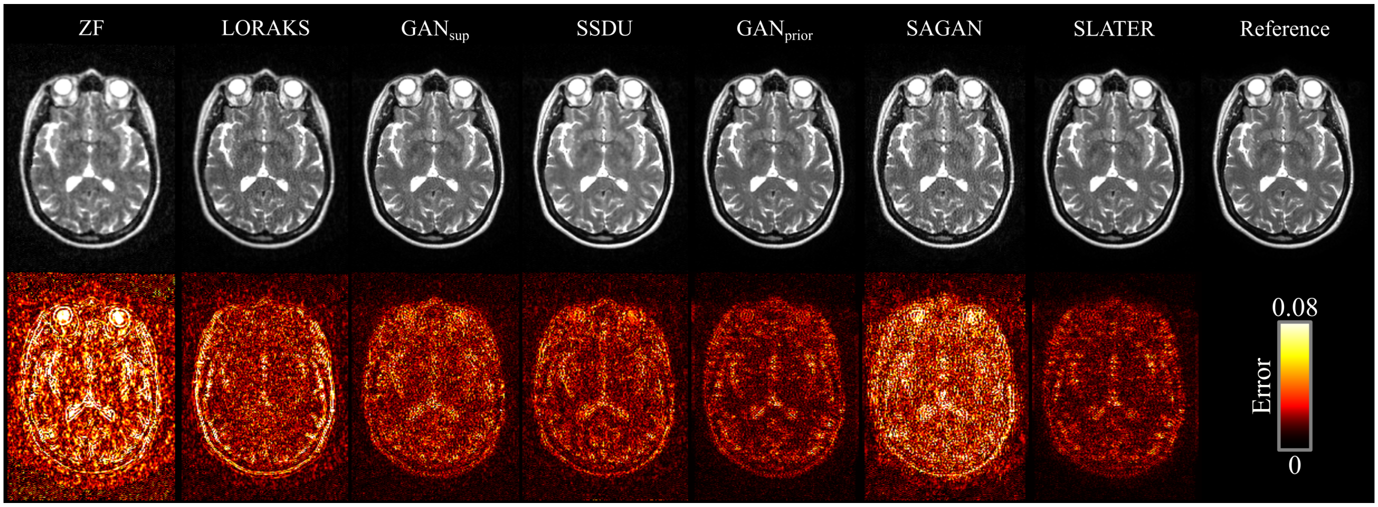
Supp. Fig. 6: Within-domain reconstructions of a T_1-weighted acquisition in the IXI dataset at R=4 are shown for the Fourier method (ZF), a traditional low-rank method (LORAKS), a supervised baseline (GAN$_{sup}$), unsupervised baselines (SSDU, GAN$_{prior}$, SAGAN) and the proposed method (SLATER) along with the reference image. Corresponding error maps are underneath the images for each method.



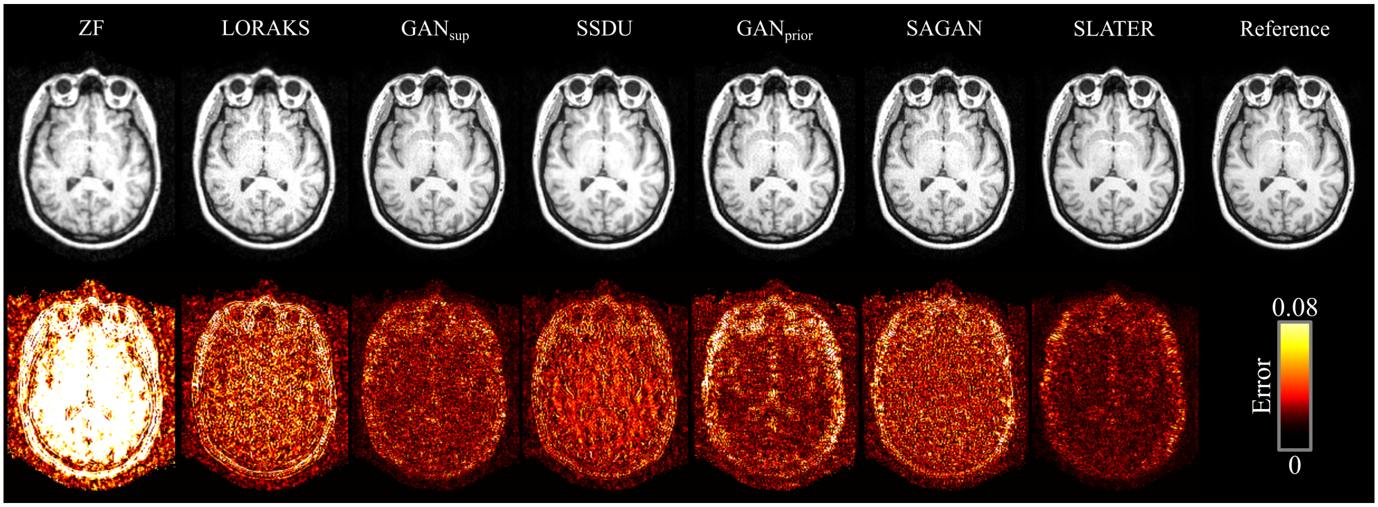
Supp. Fig. 7: Within-domain reconstructions of a T$_2$-weighted acquisition in the IXI dataset at R=4. Results are shown for ZF, LORAKS, GAN$_{sup}$, SSDU, GAN$_{prior}$, SAGAN and SLATER along with the reference image. Corresponding error maps are underneath the images for each method.



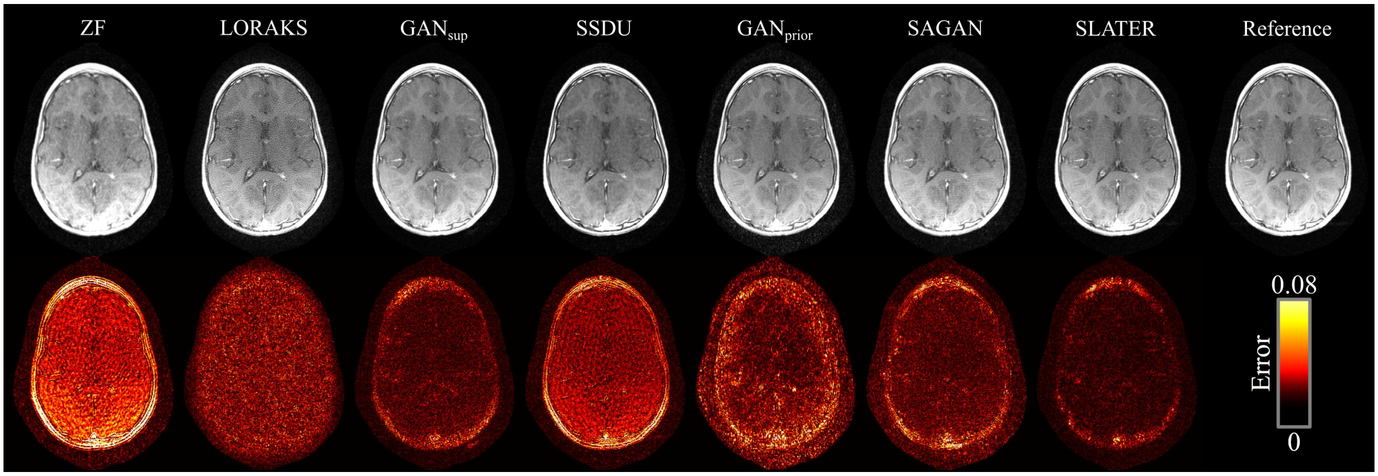
Supp. Fig. 8: Within-domain reconstructions of a T_2-weighted acquisition in the fastMRI dataset at R=4. Results are shown for ZF, LORAKS, GAN$_{sup}$, SSDU, GAN$_{prior}$, SAGAN and SLATER along with the reference image. Corresponding error maps are underneath the images for each method.



Supp. Fig. 9: Across-domain reconstructions of a T_2-weighted acquisition in the IXI dataset at R=4. Results are shown for ZF, LORAKS, GAN$_{sup}$, SSDU, GAN$_{prior}$, SAGAN and SLATER along with the reference image. Corresponding error maps are underneath the images for each method.



Supp. Fig. 10: Across-domain reconstructions of a T_1-weighted acquisition in the IXI dataset at R=4. Results are shown for ZF, LORAKS, GAN$_{sup}$, SSDU, GAN$_{prior}$, SAGAN and SLATER along with the reference image. Corresponding error maps are underneath the images for each method.



Supp. Fig. 11: Across-domain reconstructions of a T_1-weighted acquisition in the fastMRI dataset at R=4. Results are shown for ZF, LORAKS, GAN$_{sup}$, SSDU, GAN$_{prior}$, SAGAN and SLATER along with the reference image. Corresponding error maps are underneath the images for each method.