AeSPa : Attention-guided Self-supervised Parallel imaging for MRI Reconstruction

Supplementary Material

Overview. This supplementary material provides detailed insights into the implementation and data processing methodologies of our proposed AeSPa for accelerated MRI reconstruction. Sec. 1 presents quantitative comparisons with baseline methods, emphasizing trade-offs in computational time and reconstruction quality. Sec. 2 highlights qualitative results on FastMRI knee and brain datasets, showcasing AeSPa's robustness across sampling strategies and reduction factors. Sec. 3 covers implementation details, including model configurations, training setups, and data processing techniques such as normalization and filtering for standardized evaluation.

1. Additional Quantitative Results

R = 8 / Gaus.	A5000 × 1 (reported)		A6000 Ada × 1		
Model (iter)	AesPA (2000)	SPICER	AesPA(2000)	SPICER	AesPA(300)
Time	12m	10m	7m 8s	5m 12s	**1m 9s**
PSNR↑	32.57	29.50	32.57	29.50	30.01

Table 1. **Reconstruction time for various settings.**

R = 4 / Gaus. (Knee)	PSNR↑	SSIM↑
AeSPa	35.99	0.899
w/o DSS	33.02	0.837
w/o cross-channel constraint	34.14	0.825
w/o DSS & cross-channel constraint	32.11	0.823

Table 2. **Ablation study on DSS and cross-channel constraint.**

Setting	R = 4 / Uni.	R = 8 / Uni.	R = 4 / Gaus.	R = 8 / Gaus.
AeSPa	34.23 (+0.65)	28.37 (+1.18)	35.99 (+0.29)	32.57 (+0.25)
w/o AKSM	33.58	27.19	35.70	32.32

Table 3. **AKSM ablation study on the FastMRI Knee dataet.**

masking ratio	PSNR ↑	SSIM ↑
70%	**32.57 ± 3.24**	**0.853 ± 0.05**
50%	32.32 ± 3.29	0.852 ± 0.05
30%	32.33 ± 3.26	0.853 ± 0.05
10%	32.07 ± 3.71	0.843 ± 0.06
5%	31.59 ± 3.79	0.836 ± 0.07
0%	29.91 ± 4.48	0.806 ± 0.09

Table 4. **Ablation study results on the masking ratio of k-space in the training process for the FastMRI knee dataset using 1D Gaussian sampling with R = 8.**

Table 1 presents the comparison of reconstruction time and performance across various settings. This table compares the execution time of the AesPA and SPICER models using two different GPUs (A5000 and A6000 Ada) while evaluating the reconstruction quality in terms of PSNR. As reported in our manuscript, on the A5000 GPU, AesPA requires 12 minutes, while SPICER takes 10 minutes. When using the more advanced A6000 Ada GPU, the overall speed improves, with AesPA completing in 7 minutes and 8 seconds and SPICER in 5 minutes and 12 seconds. These results are based on 2000 iterations, targeting the best possible performance. However, by optimizing the number of iterations, AesPA can achieve a reconstruction time of just 1 minute and 9 seconds with only 300 iterations, demonstrating that reconstruction can be completed in approximately 1 minute per slice.

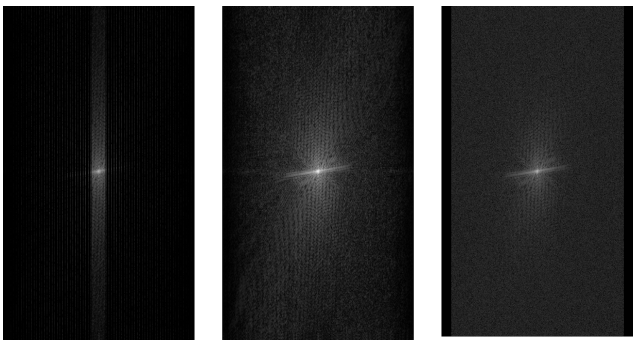

Figure 1. **Qualitative results in k-space.** The first column shows the undersampled k-space (1D Uniform, R = 4), the second column depicts $\mathbf{k}_{sel}$, and the third column represents the fully sampled k-space.

Our cross-channel constraint differs conceptually from DSS: considering that each coil image is the product of the true MR image and its spatially varying sensitivity map, the RSS (Root Sum of Squares) of sensitivity maps across channels should theoretically yield a ones matrix. Our constraint enforces this physical property at each pixel position, while DSS simply normalizes the maximum value of each coil channel to 1 (channel-wise). This physically-motivated cross-channel constraint strongly regularizes sensitivity estimates, resulting in a more accurate sensitivity map. As shown in Table 2, the cross-channel constraint significantly improves reconstruction performance over DSS alone.

As shown in Table 3, AKSM improves PSNR by up to **+1.18dB**. AKSM selectively determines important k-space in non-sampled line, its impact was slightly limited in Gaussian sampling where significant k-space lines are already acquired, resulting in modest PSNR improvement.

An ablation study on the k-space masking ratio in IKM,

summarized in Table 4, demonstrates that a 70% masking ratio achieves optimal performance, highlighting the efficacy of selective k-space data reduction in preserving reconstruction fidelity.

	Metric	Uniform1d		Gaussian1d	
		R = 4	R = 8	R = 4	R = 8
E2E-VarNet	PSNR↑	34.89 ± 2.25	31.04 ± 2.66	35.01 ± 2.06	31.76 ± 2.22
	SSIM↑	0.944 ± 0.01	0.893 ± 0.02	0.947 ± 0.01	0.902 ± 0.02
Ours	PSNR↑	33.37 ± 4.01	27.54 ± 3.60	33.98 ± 4.16	28.47 ± 3.39
	SSIM↑	0.889 ± 0.05	0.789 ± 0.07	0.899 ± 0.06	0.801 ± 0.07

Table 5. **Quantitative results of our methods with supervised learning method.**

As presented in Table 5, we compared our method to pretrained E2E-VarNet, a widely used supervised learning model for MR reconstruction. To ensure a fair comparison with the pretrained model, we applied the preprocessing method used in FastMRI to our method during the experiments. While our approach showed slightly lower performance in terms of PSNR and SSIM compared to E2E-VarNet, it is important to note that E2E-VarNet relies on extensive fully-sampled training data, which is often challenging to obtain in real clinical settings. In contrast, our method offers a distinct advantage with its ability to perform scan-specific reconstruction without the need for pretrained datasets, underscoring its practical utility.

2. Additional Qualitative Results

Figure 1 illustrates the qualitative results of k-space for different sampling strategies. The first column shows the undersampled k-space obtained using 1D Uniform sampling with a reduction factor of R = 4. The second column depicts $\mathbf{k}_{sel}$, which represents the selectively reconstructed k-space. The third column displays the fully sampled k-space as the ground truth. This comparison highlights the effectiveness of $\mathbf{k}_{sel}$ in bridging the gap between undersampled and fully sampled k-space, preserving essential structural details.

Figures 2, 3 and 4 illustrate the qualitative results of accelerated MRI reconstruction on FastMRI knee and brain datasets using different sampling strategies and reduction factors. Figure 2 compares methods at a reduction factor of R = 4 with 1D Gaussian random sampling for knee MRI, where AeSPa (Ours) achieves the highest PSNR (33.61) and SSIM (0.907), demonstrating sharper details and fewer artifacts compared to other methods. Figure 3 presents similar results for R = 8, showing AeSPa's consistent superiority in preserving fine structures with PSNR (34.09) and SSIM (0.899). Figure 4 shifts focus to brain MRI with R = 4 using 1D Uniform sampling, where AeSPa outperforms all baselines with a PSNR of 36.98 and SSIM of 0.964, achieving clear edges and minimal noise. These results highlight AeSPa's ability to reconstruct high-quality images across various anatomical regions and sampling conditions, outperforming both traditional and learning-based approaches. Figure 5 shows the qualitative performance of our model with reduction factors ranging from R = 8 to R = 23 using 1D Gaussian random sampling. At R = 8, the model achieves the highest PSNR (40.09 dB) and SSIM (0.974). As the reduction factor increases, reconstruction quality gradually decreases but remains robust, with PSNR and SSIM stabilizing at 34.42 dB and 0.936 for R = 23. This demonstrates the model's resilience under extreme undersampling.

3. Details

3.1. Implementations

In this paper, the experiments were conducted using an NVIDIA RTX A5000, and the code was implemented based on PyTorch. The coil combined image and Sensitivity Map estimation model were implemented using U-Net as the foundation, with the Mamba module added between the encoder and decoder. All models processed complex data by splitting it into real and imaginary parts, using an input and output format of (channel, height, width) $\times$ 2. The first part was treated as the real part and the second as the imaginary part for model training. The learning rates for coil combined image, the Sensitivity Map estimation model, and AKSM were set to 5e-4, 5e-4, and 5e-5, respectively, with the Adam optimizer. For training, the raw data from the dataset was used after basic preprocessing as mentioned in the data preprocessing section. For model evaluation, the data was center cropped to 320×320 size, followed by max normalization and a bandpass filter to assess model performance. The Mamba model parameters used in AKSM were dmodel = 640 (kspace height), dstate = 64, dconv = 640 (kspace height), and expand = 1. The Mamba model parameters for the U-Net were dmodel = 1024 (latent vector), dstate = 16, dconv = 4, and expand = 2. Layer normalization was applied after every Mamba module. In AKSM, to use k-space as input to the Mamba layer, the data is flattened from the form $\mathbb{R}^{kx \times ky \times C \times 2}$ into a 1D sequence of shape $\mathbf{L} \in \mathbb{R}^{B \times (C \times 2 \times ky) \times kx}$, sequentially aligned along the *ky* direction. Here, C represents the number of coils, and 2 denotes the real and imaginary channels. The batch size was 1, and the models were trained for 2,000 iterations ($\mathbf{T}$), with the inputs of the coil combined image estimation model and Sensitivity Map Estimator updated every 1 and 10 iterations, respectively.

3.2. Data Processing

We normalized the k-space data by dividing it by the maximum value of its corresponding coil-combined image, which was obtained through inverse Fourier transform followed by coil combination. This normalization strategy

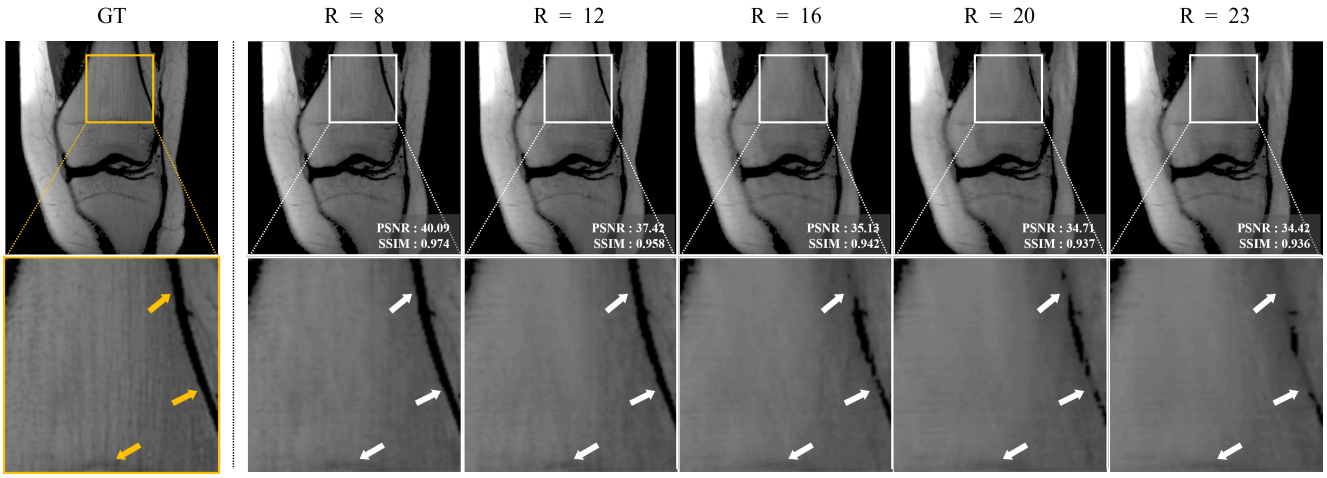

Figure 2. **Qualitative results for accelerated MRI reconstruction on the FastMRI knee data**. Reconstruction was performed using 1D Gaussian random sampling with R = 4.

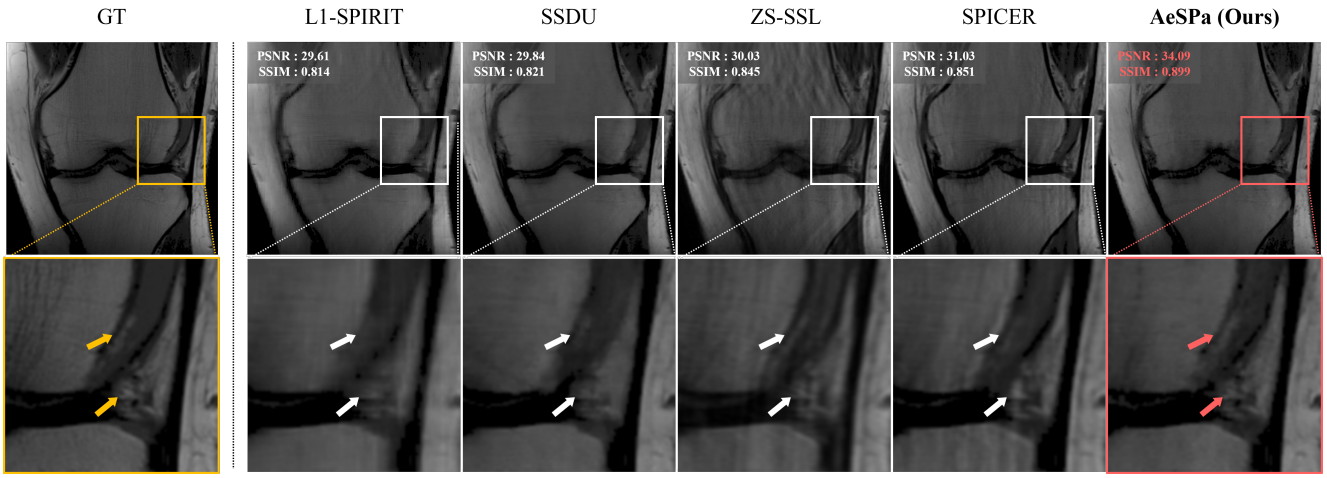

Figure 3. **Qualitative results for accelerated MRI reconstruction on the FastMRI knee data**. Reconstruction was performed using 1D Gaussian random sampling with R = 8.

ensures that the image domain values approximately fall within the range of [0, 1]. During evaluation, all coil-combined images from both comparison models and our model were additionally processed with max normalization to maintain the value range within [0, 1]. To remove noise outside the FOV region, a bandpass filter was applied to both the ground truth and reconstructed images (knee: [0.7, 130], brain: [0.8, 200]).

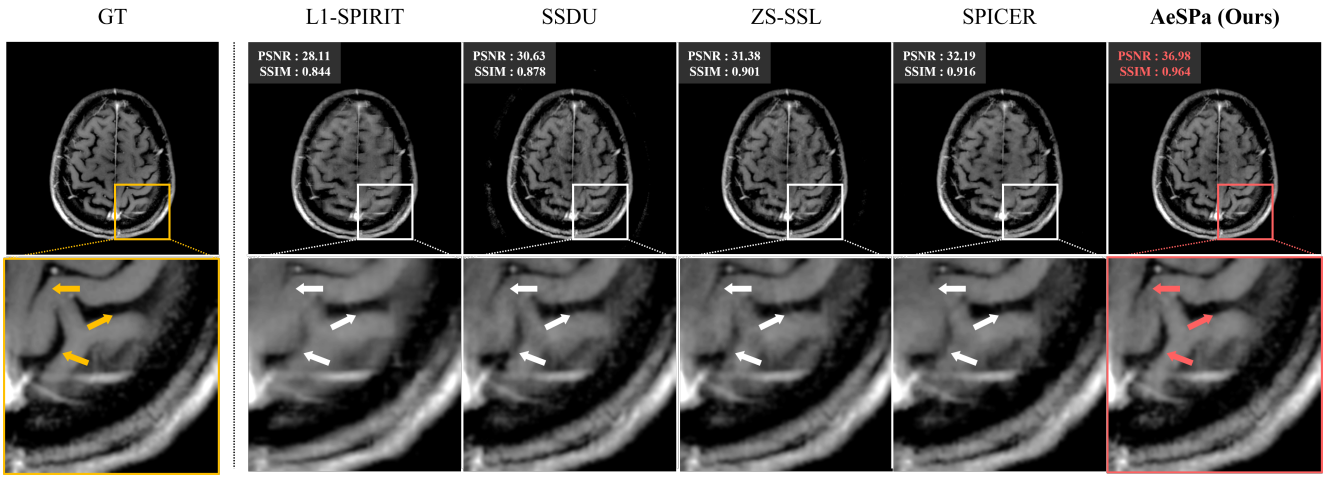

Figure 4. **Qualitative results for accelerated MRI reconstruction on the FastMRI brain data**. Reconstruction was performed using 1D Uniform sampling with R = 4.

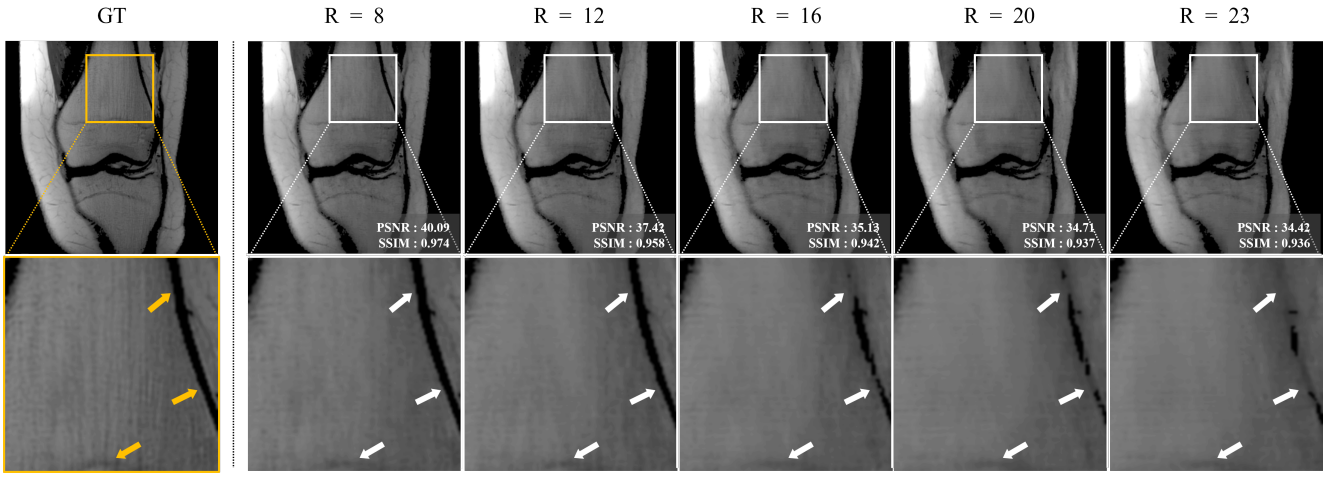

Figure 5. **Qualitative results of our model with varying reduction factors**. The performance is demonstrated for 1D Gaussian random sampling with reduction factors ranging from R = 8 to 23.