Adaptive Anisotropic Gaussian Splatting for Multi-contrast MRI Arbitrary-Scale Super-Resolution with Anatomy Guidance

Supplementary Material

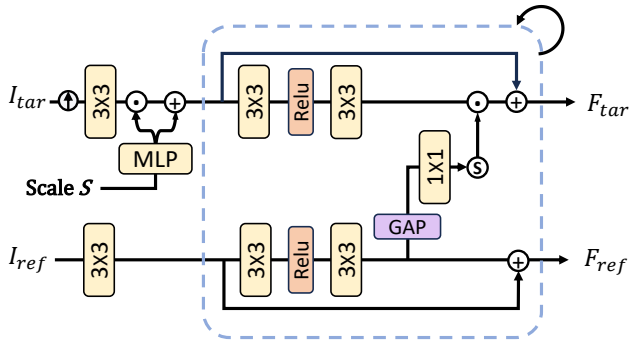


Figure 8. Network architecture of the CSMF Encoder

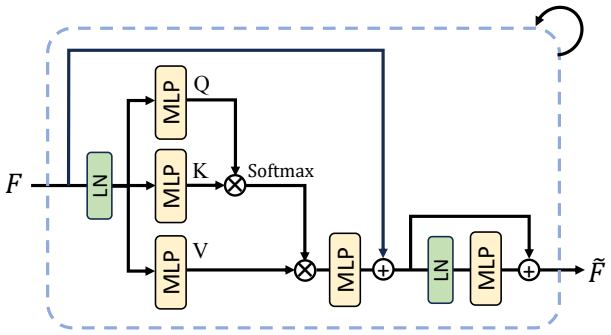


Figure 9. Network architecture of the Gaussian Transformer

8. Supplementary Network Architecture

8.1. Cross-Scale Multi-contrast Fusion Encoder

The Cross-Scale Multi-contrast Fusion (CSMF) encoder, depicted in Fig. 8, is designed to extract and fuse multi-scale deep features from the reference image I_{ref} and target image I_{tar}. We first align the spatial resolution of I_{tar} to I_{ref} via k-space zero-padding resampling. Both images are then passed through three 3×3 convolutional layers (with ReLU) to extract initial features. To incorporate scale awareness, the scale factor S is embedded via an MLP and infused into the target features through element-wise multiplication and addition. For cross-contrast fusion, the reference features are globally averaged and processed by a 1×1 convolution to produce a channel attention vector, which adaptively reweights the target features channel by channel. Finally, both branches are enhanced with a residual path consisting of 16 stacked residual units to boost the encoder's representational capacity.

8.2. Gaussian Transformer

As illustrated in Fig. 9, the Gaussian Transformer refines the input Gaussian feature map F via a self-attention mechanism. The process begins by applying layer normalization to F, followed by three MLP projections to generate the query (Q), key (K), and value (V) tensors. The attention weights are computed as $\mathrm{softmax}(QK^\top)$, which are used to aggregate the value features. This is followed by a feed-forward MLP with a residual connection and a second pre-normalization step, producing the refined representation $\tilde{F}$. By stacking five such blocks, the model enables comprehensive global interactions among all Gaussians, significantly enhancing the coherence of the final representation.

9. Runtime Efficiency

We evaluate the runtime efficiency of state-of-the-art arbitrary-scale MR image SR methods under identical hardware and testing protocols. As shown in Table 3, GaussM^{2}ASR requires only 2.14/2.11 GB GPU memory and 320/287 ms per image at $4\times/6\times$, respectively. Compared with DINet, our method reduces GPU memory from 20.13/19.53 GB to 2.14/2.11 GB and inference time from 681/657 ms to 320/287 ms. Compared with GSASR, our method also achieves substantially lower memory consumption (6.61/3.49 GB vs. 2.14/2.11 GB). Meanwhile, even compared with the lightweight dual-arb baseline, GaussM^{2}ASR reduces GPU memory usage by 23.8% at both scales while maintaining practical inference speed. These results demonstrate that GaussM^{2}ASR offers deployment-friendly efficiency for clinical applications.

10. Additional ablation experiments

10.1. Robustness to Cross-Contrast Differences and HR–LR Misalignment

We further evaluate the robustness of GaussM^{2}ASR to cross-contrast discrepancies and residual HR–LR misalignment, with results summarized in Table 4. For the cross-contrast setting, we use PD images as references for T2 super-resolution while the model is trained on T1–T2 pairs. Despite the substantial appearance gap, our method achieves the best PD-T2 performance, surpassing DINet by 1.25 dB PSNR and 0.0341 SSIM. We also assess robustness to residual misalignment using synthetic non-rigid perturbations. In this setting, GaussM^{2}ASR again obtains the best results, outperforming DINet by 0.99 dB PSNR and 0.0260 SSIM. These results demonstrate that our anatomy-guided

Table 3. Efficiency comparison with state-of-the-art arbitrary-scale MRI image SR methods on the IXI dataset at $4\times$ and $6\times$ scaling factors. All methods are benchmarked under identical hardware and testing protocols. The reference image and the GT HR target image both have a resolution of 256×256.

Method	Params (M)	$4\times$		$6\times$	
		GPU Mem. (GB)	Inference Time (ms)	GPU Mem. (GB)	Inference Time (ms)
GaussianSR	1.80	1.51	28.7	1.44	27.8
GSASR	19.13	6.61	139.3	3.49	97.8
dual-arb	13.75	2.81	85.5	2.77	79.4
McASSR	3.20	1.15	123	1.15	111.8
DINet	26.00	20.13	681	19.53	657
Ours	17.49	2.14	320	2.11	287

Table 4. Robustness to cross-contrast differences and residual HR–LR misalignment on the IXI dataset at a $4\times$ scaling factor. In all experiments, the HR target image has a resolution of 256×256. For the PD-T2 setting, PD images are used as references for T2 super-resolution.

Method	Non-rigid		PD-T2	
	PSNR	SSIM	PSNR	SSIM
GaussianSR	28.33	0.8698	26.38	0.8429
GSASR	28.51	0.8711	26.54	0.8457
LIIF	28.84	0.8745	26.95	0.8487
Dual-ArbNet	29.47	0.8839	27.48	0.8588
McASSR	29.88	0.8969	27.57	0.8667
DINet	30.44	0.9033	28.34	0.8723
Ours	**31.43**	**0.9293**	**29.59**	**0.9064**

Table 5. Edge-aware evaluation and downstream segmentation results on the IXI dataset at a $4\times$ scaling factor. In all experiments, the HR target image has a resolution of 256×256. Lower RMSE is better, while higher F-score, GMC, and Dice indicate better structural fidelity.

Model	RMSE↓	F-score↑	GMC↑	Dice↑
GaussianSR	10.791	0.919	0.720	0.796
GSASR	8.581	0.944	0.768	0.857
LIIF	9.318	0.934	0.735	0.840
Dual-ArbNet	8.355	0.937	0.733	0.862
McASSR	8.161	0.948	0.749	0.886
DINet	7.472	0.952	0.761	0.871
Ours	**6.659**	**0.956**	**0.773**	**0.932**

Gaussian representation can effectively suppress unreliable reference cues and adapt to local spatial discrepancies, enabling faithful high-frequency recovery without introducing spurious boundary artifacts. This validates the necessity of our anatomy-aware matching strategy over naive reference transfer.

Table 6. Ablation study on the number of Gaussians in GaussM^{2}ASR on the IXI dataset at a $4\times$ scaling factor.

Number	PSNR	SSIM	GPU Mem. (GB)	Time (ms)
baseline	32.03	0.9350	2.14	320
$4\times$ baseline	32.09	0.9353	4.95	718.8
$9\times$ baseline	31.96	0.9347	7.56	897.7

10.2. Additional Edge-Aware Metrics

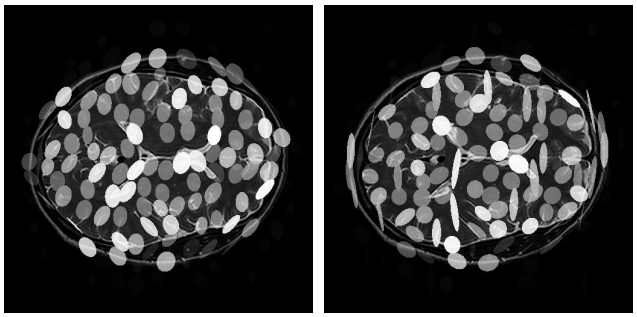
We additionally report edge-aware metrics and downstream segmentation results, with results summarized in Table 5. Compared with existing methods, GaussM^{2}ASR achieves the lowest Edge RMSE and the highest Boundary F-score, GMC, and Dice. In particular, relative to DINet, our method reduces RMSE from 7.472 to 6.659 while improving the F-score from 0.952 to 0.956, GMC from 0.761 to 0.773, and Dice from 0.871 to 0.932. These results indicate that the recovered high-frequency details are not only visually sharper but also more anatomically consistent, leading to more accurate boundary localization and improved downstream clinical analysis.

10.3. Effect of the Number of Gaussians

We conduct a Gaussian-number ablation study to assess the impact of model capacity, with results summarized in Table 6. Compared to the baseline, increasing the number of Gaussians to $4\times$ yields only marginal performance gains, while a $9\times$ increase leads to a noticeable performance drop. Moreover, both expanded configurations incur substantial computational overhead, significantly increasing inference time and GPU memory usage. These results demonstrate that simply increasing Gaussian count fails to deliver proportional performance improvements and can even introduce optimization instability. This underscores the necessity of our anatomy-aware Gaussian allocation strategy over naive quantity expansion.

Table 7. Encoder ablation study of GaussM^{2}ASR on the IXI dataset at a $4\times$ scaling factor.

Method	PSNR	SSIM
RDN Encoder	31.96	0.9347
Ours	32.03	0.9350



(a) Without Top-T (b) With Top-T

Figure 10. Visualization of Gaussian ellipse distributions with and without the Top-T module

10.4. Effect of the Encoder

To evaluate the encoder design, we replace the proposed CSMF encoder with a conventional backbone using two independent Residual Dense Networks (RDNs) for feature extraction. As reported in Table 7, this substitution results in a slight performance drop compared to the full GaussM^{2}ASR model with the CSMF encoder. Nevertheless, this RDN-based variant still significantly outperforms other INR-based multi-contrast MRI SR methods that also employ RDN backbones. This ablation study confirms that while the CSMF encoder provides specialized benefits for cross-scale and multi-contrast fusion, the core advantage of GaussM^{2}ASR stems from its Gaussian representation, which is inherently more powerful than INR-based approaches for modeling HF anatomical details.

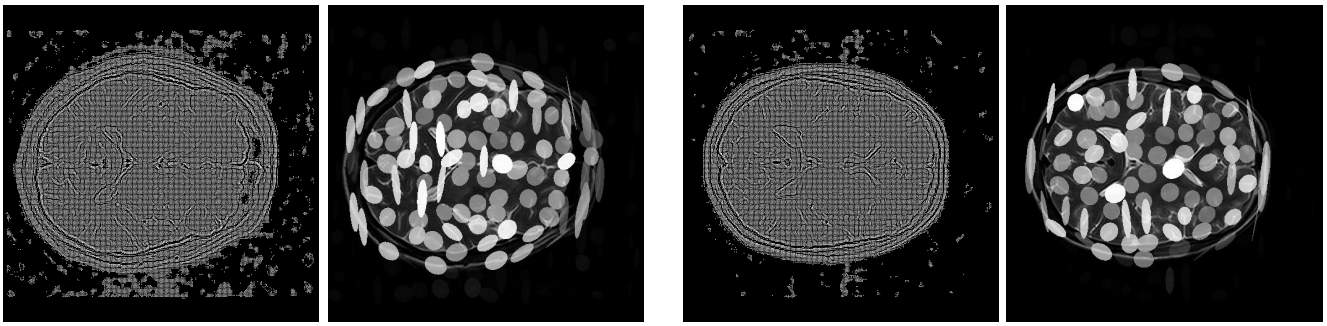
10.5. Additional Analysis of the Top-T Module

To intuitively analyze the impact of the Top-T module, we visualize the Gaussian ellipse distributions with and without it in Fig. 10. The introduction of Top-T enables the Gaussian centers to concentrate along anatomical boundaries and the ellipses to adaptively narrow in HF regions, thereby achieving more accurate structural fitting and richer HF detail.

11. Additional Visual Results

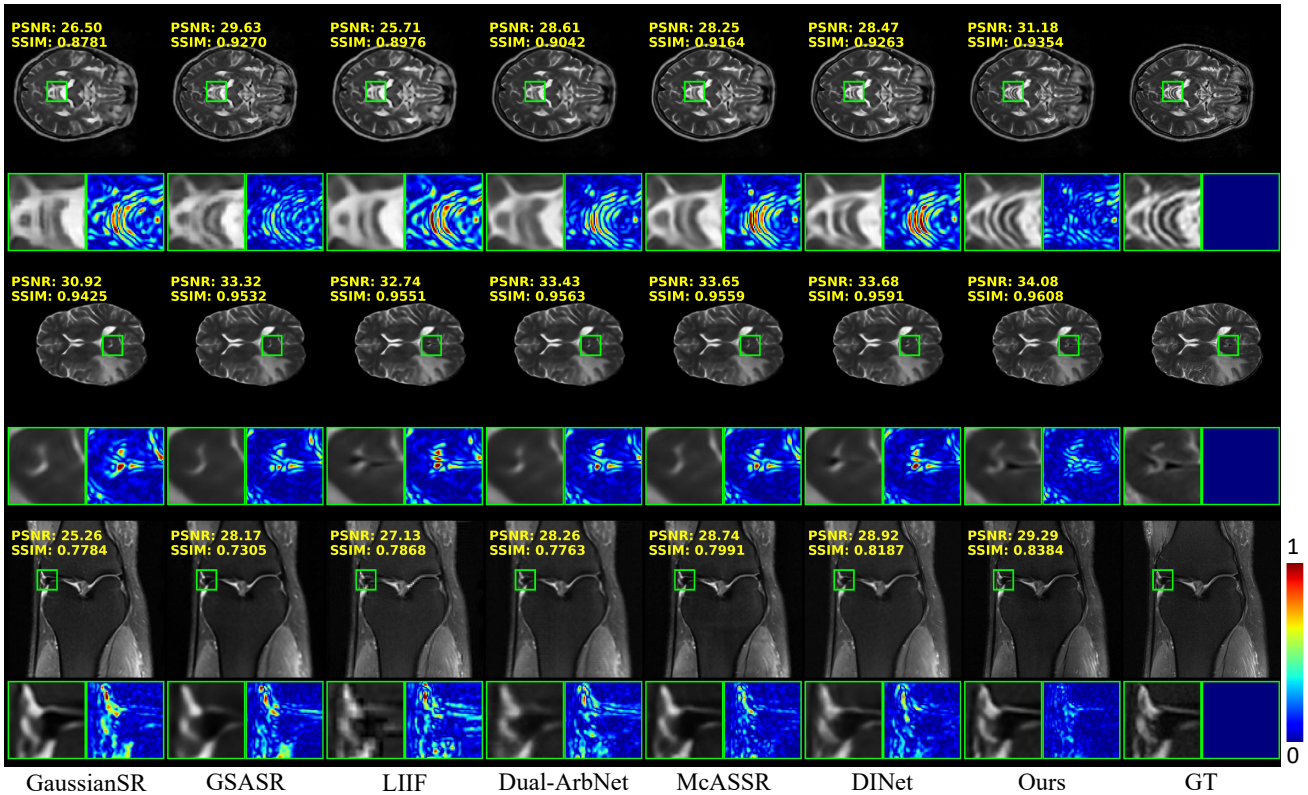
To analyze the learned Gaussian distributions, we employ two complementary visualization schemes in Fig. 11. First, we generate a spatial density map by projecting the normalized coordinates of each Gaussian center back to the image plane and plotting them as dots, reflecting the overall distribution density. Second, we create an anisotropic coverage map by rendering elliptical regions onto the original slice, where each ellipse's size, shape, and intensity are determined by the Gaussian's covariance and opacity ($\boldsymbol{\alpha}$). This scheme intuitively reveals the model's attention to different regions and their anisotropic characteristics.

Finally, we present additional visualizations of the experimental results, as shown in Figs. 12–15.



Position Distribution | Visual Gaussians | Position Distribution | Visual Gaussians

Figure 11. More visualizations of Gaussian positions and Gaussian Ellipse distributions



GaussianSR | GSASR | LIIF | Dual-ArbNet | McASSR | DINet | Ours | GT

Figure 12. Qualitative comparison on IXI, BraTS, and fastMRI datasets at a 4× scaling factor. Enlarged views of regions within green boxes are provided alongside their error maps, which display the absolute reconstruction error (darker shades indicate smaller errors).

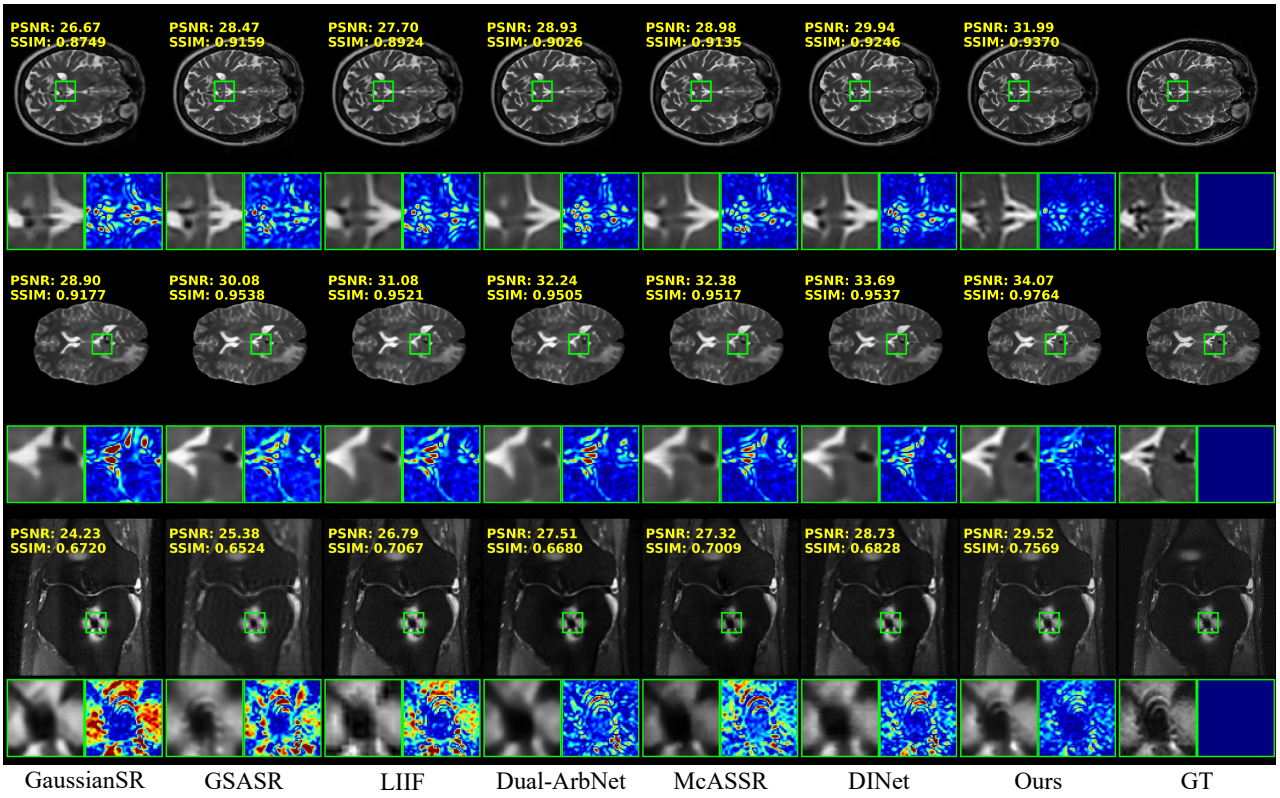


Figure 13. Qualitative comparison on IXI, BraTS, and fastMRI datasets at a 4× scaling factor. Enlarged views of regions within green boxes are provided alongside their error maps, which display the absolute reconstruction error (darker shades indicate smaller errors).

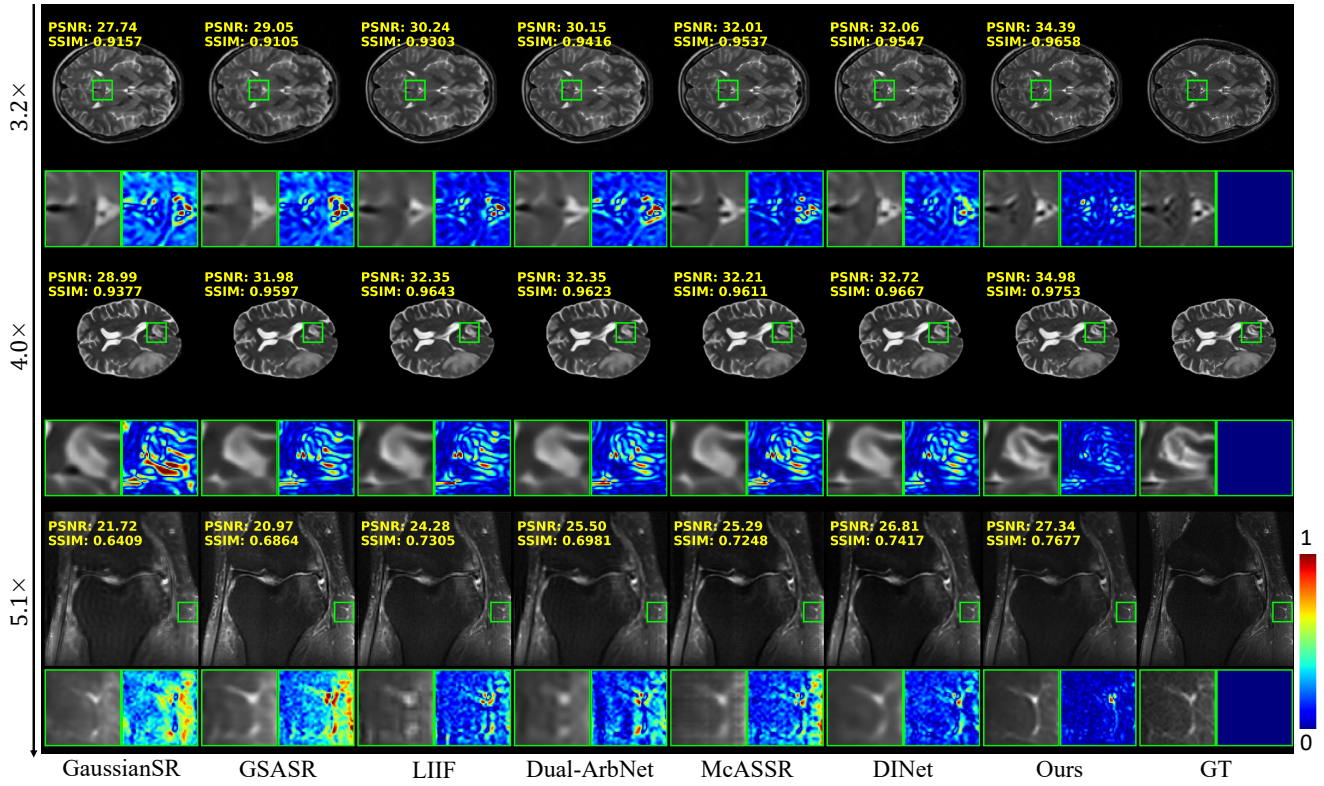


Figure 14. Qualitative comparison on IXI, BraTS, and fastMRI datasets. Enlarged views of regions within green boxes are provided alongside their error maps, which display the absolute reconstruction error (darker shades indicate smaller errors).

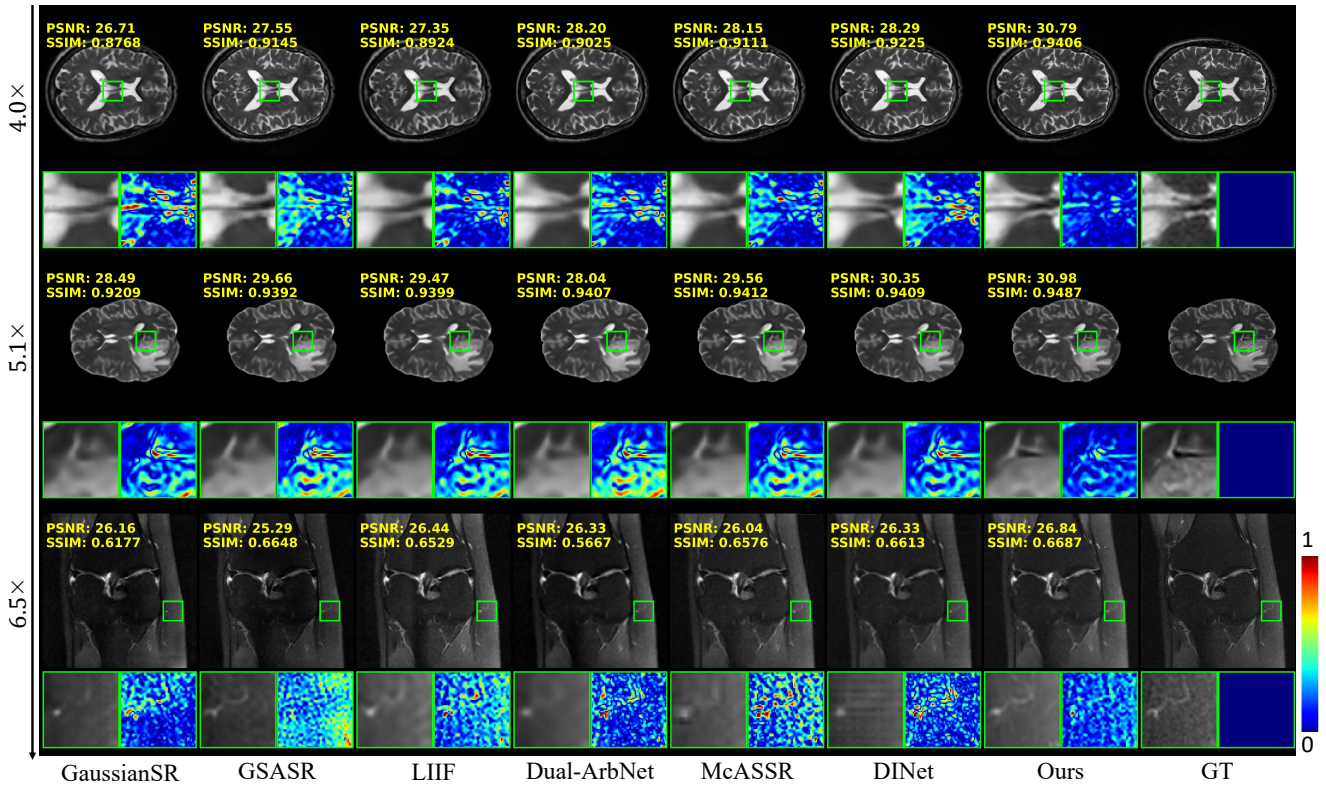


Figure 15. Qualitative comparison on IXI, BraTS, and fastMRI datasets. Enlarged views of regions within green boxes are provided alongside their error maps, which display the absolute reconstruction error (darker shades indicate smaller errors).