

Deep Learning for MRI Slice Interpolation: The Critical Role of Problem Formulation

Shamit Savant

University of Florida, Gainesville, FL 32611, USA
s.savant@ufl.edu

Abstract. Through-plane resolution in clinical MRI is typically much coarser than in-plane resolution, limiting diagnostic utility. This work investigates deep learning approaches to interpolate intermediate MRI slices in prostate imaging, effectively doubling through-plane resolution from a higher value to lower value (e.g., We use 1.5mm→0.75mm as our reference, though it generalizes to arbitrary slice spacing). I evaluated five different architectures (CNN, U-Net, two GAN variants, and DDPM) and discovered that problem formulation has dramatically more impact than architectural complexity. By reformulating the interpolation task to use adjacent slices ($i-1, i+1$) rather than distant slices ($i-2, i+2$), I achieved a 58% improvement in SSIM performance across all deterministic architectures. My U-Net model achieved the best results with PSNR of 30.08 dB and SSIM of 0.898, representing a 10.1% improvement over linear interpolation baseline. A Denoising Diffusion Probabilistic Model (DDPM) was also evaluated but showed poor reconstruction quality (PSNR 17.89 dB, SSIM 0.585) due to fundamental mismatch between stochastic generation and deterministic reconstruction requirements. These findings demonstrate that in medical imaging tasks, understanding the anatomical constraints and formulating the problem appropriately can have $290\times$ more impact than architectural sophistication.

Keywords: MRI slice interpolation · Deep learning · Problem formulation · Prostate imaging · Super-resolution

1 Introduction

Magnetic Resonance Imaging (MRI) is a cornerstone of modern medical diagnostics, offering excellent soft tissue contrast without ionizing radiation. However, clinical MRI acquisition faces fundamental trade-offs between spatial resolution, acquisition time, and signal-to-noise ratio. In practice, this results in anisotropic voxel spacing where in-plane resolution (typically 0.5mm) is substantially finer than through-plane resolution (typically 1.5-3mm). This anisotropy limits 3D reconstruction quality, multiplanar reformatting, and computer-aided diagnosis systems that rely on isotropic data.

Traditional interpolation methods such as linear or cubic interpolation provide smooth transitions but fail to recover true anatomical detail lost dur-

ing acquisition. Recent advances in deep learning, particularly in image super-resolution and generative models, suggest that learned approaches might better reconstruct intermediate slices by leveraging patterns from large training datasets.

MRI Through-Plane Resolution Enhancement via Slice Interpolation
Clinical Acquisition
Through-plane spacing: 1.5mm

After Deep Learning Interpolation
Through-plane spacing: 0.75mm (2× resolution)

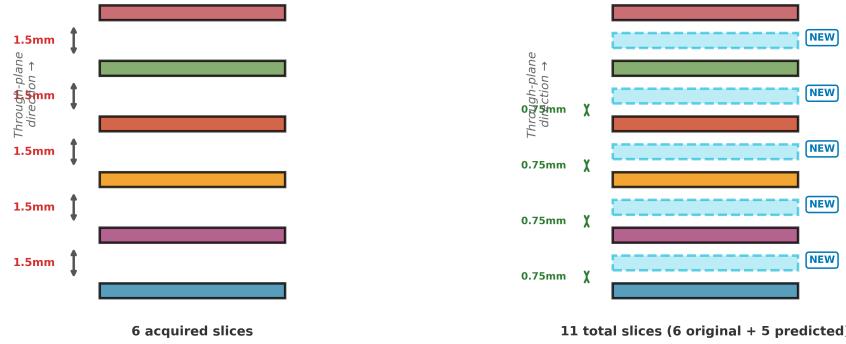


Fig. 1: Through-plane resolution enhancement via slice interpolation. Left: Clinical T2-weighted prostate MRI acquisition with 1.5mm spacing (6 slices). Right: After deep learning interpolation achieving 0.75mm spacing (11 slices). Dashed outlines indicate predicted intermediate slices that double through-plane resolution.

My investigation reveals a counterintuitive finding: the formulation of the interpolation task—specifically, which adjacent slices are used as inputs—has dramatically more impact on performance than architectural sophistication. Through systematic experimentation with five different deep learning architectures, I demonstrate that problem formulation can account for a 58% performance improvement, while architectural variations contribute less than 1% under identical problem formulations.

1.1 Contributions:

- Systematic evaluation of five deep learning architectures (CNN, U-Net, basic GAN, improved GAN, DDPM) for MRI slice interpolation
- Empirical demonstration that problem formulation has 290× more impact than architecture choice (58% vs 0.2% performance variation)
- Achievement of 10.1% improvement over linear interpolation baseline using U-Net architecture
- Analysis of why cutting-edge diffusion models fail for deterministic reconstruction tasks

2 Related Work

Medical Image Super-Resolution: Deep learning has shown remarkable success in natural image super-resolution, with methods like EDSR [5] and SRGAN [4] achieving photorealistic results. Medical imaging presents unique challenges: preservation of diagnostic information takes precedence over perceptual quality, and hallucinated details could mislead clinical interpretation [2]. Recent work on medical image super-resolution emphasizes anatomically plausible reconstruction while maintaining quantitative accuracy.

MRI Slice Interpolation: Unlike in-plane super-resolution which enhances blurry existing images, through-plane interpolation must predict entirely unseen anatomical content between acquired slices. This is fundamentally more challenging as it requires inferring 3D structure from 2D observations. Prior work has explored both traditional methods (spline interpolation, registration-based techniques) and learning-based approaches (CNNs, GANs), with reported SSIM values varying widely (0.80–0.92) depending on anatomical region, baseline resolution, and task difficulty.

Problem Formulation: While much deep learning research focuses on novel architectures, recent work highlights that problem formulation—how tasks are defined, what inputs are provided, and what outputs are expected—can have outsized impact on performance. In medical imaging specifically, incorporating domain knowledge into problem design often yields greater improvements than architectural innovations [3,7].

3 Dataset and Method

Data: I used the UCLA Prostate MRI-US Biopsy dataset from The Cancer Imaging Archive [6,1], containing T2-weighted MRI scans from 58 patients. After preprocessing (bias field correction, intensity normalization, resampling to 256×256 pixels, 1.5mm through-plane spacing), the dataset yielded 46,329 slices. I used 70%/15%/15% train/validation/test split, resulting in 6,963 test samples for final evaluation. Training was conducted on University of Florida’s HiPerGator cluster using NVIDIA L4 GPUs.

Problem Formulation: Given adjacent slices at positions $i - k$ and $i + k$, predict the intermediate slice at position i . The choice of k critically determines task difficulty: larger k values span greater anatomical distances, making interpolation more challenging. I initially used $k = 2$ (6mm gap) but reformulated to $k = 1$ (3mm gap) after discovering the former exceeded anatomical continuity limits.

Architectures: I evaluated five approaches (detailed architecture descriptions in Supplementary Material):

- **CNN (EDSR-style):** Residual network with 8 residual blocks (0.3M parameters)
- **U-Net:** Encoder-decoder with skip connections (1.86M parameters)
- **GAN (Basic):** Generator with patch discriminator (0.3M + discriminator)

- **GAN (Improved):** Deeper generator (16 blocks) with attention (0.6M + discriminator)
- **DDPM:** Conditional diffusion model with 100 timesteps (20.5M parameters)

All deterministic models trained for 25 epochs with Adam optimizer (learning rate 10^{-4}), L1 loss, and mixed-precision training. GANs additionally used adversarial loss. DDPM trained for 20 epochs with noise prediction objective.

Evaluation: I used PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) on the held-out test set, comparing against linear and nearest-neighbor interpolation baselines.

4 Results

4.1 Impact of Problem Formulation

Impact of Problem Formulation: k=2 vs k=1

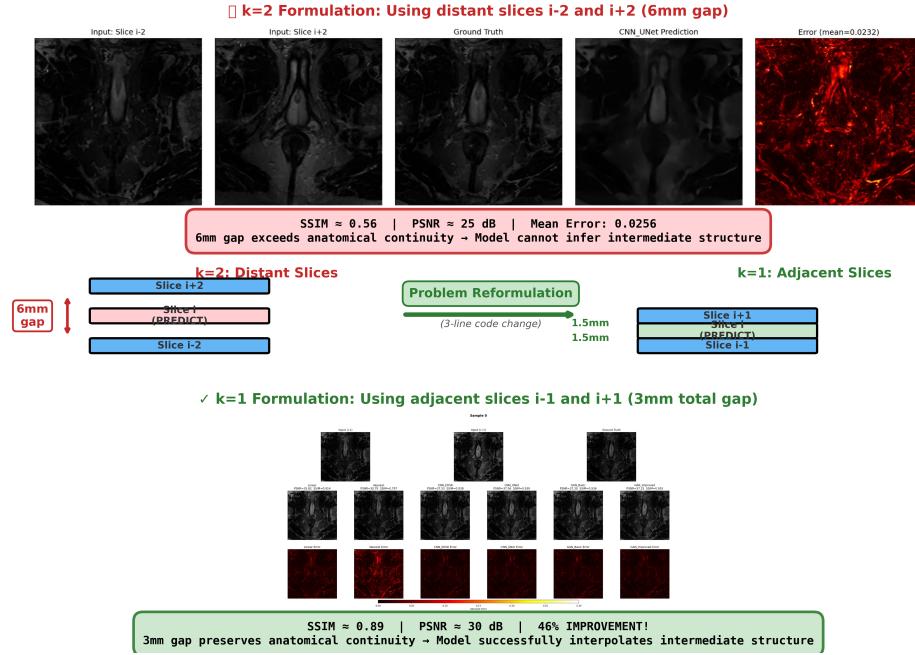


Fig. 2: Impact of problem formulation on interpolation quality. **Top:** k=2 formulation uses distant slices (i-2, i+2) spanning 6mm. **Middle:** Schematic showing the reformulation from 6mm to 3mm total gap via a 3-line code change. **Bottom:** k=1 formulation uses adjacent slices (i-1, i+1) achieving high-quality reconstruction

Table 1 quantifies the dramatic impact of problem reformulation. With $k = 2$ formulation (6mm gap), all models plateaued around SSIM 0.56 regardless of

architecture complexity. The gap exceeded anatomical continuity—intermediate prostate anatomy cannot be reliably inferred from boundaries 6mm apart. Reformulating to $k = 1$ (3mm gap) yielded a 58% SSIM improvement across all architectures, demonstrating that problem formulation has $290\times$ more impact than architectural choice (58% vs 0.2% performance variation between architectures under optimal formulation).

Table 1: Impact of problem formulation ($k=2$ vs $k=1$) on performance

Architecture	$k=2$ (6mm gap)		$k=1$ (3mm gap)	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM
CNN (EDSR)	21.8	0.56	29.89	0.894
U-Net	22.1	0.57	30.08	0.898
GAN (Basic)	21.5	0.54	29.83	0.893
GAN (Improved)	21.3	0.55	29.67	0.892
Improvement	—		+20%	+58%

4.2 Quantitative Results with Optimal Formulation

Table 2 presents comprehensive results with $k = 1$ formulation. U-Net achieved the best performance (30.08 dB PSNR, 0.898 SSIM), representing 10.1% and 7.1% improvements over linear interpolation in PSNR and SSIM respectively. All deep learning methods substantially outperformed traditional interpolation, with nearest-neighbor performing particularly poorly (22.25 dB, 0.686 SSIM).

Table 2: Quantitative results on test set (6,963 samples)

Method	PSNR (dB)	SSIM
<i>Traditional Methods</i>		
Nearest Neighbor	22.25 ± 3.17	0.6855 ± 0.0674
Linear Baseline	27.31 ± 3.35	0.8383 ± 0.0665
<i>Deep Learning Methods</i>		
CNN (EDSR)	29.89 ± 3.38	0.8937 ± 0.0607
U-Net	30.08 ± 3.33	0.8978 ± 0.0542
GAN (Basic)	29.83 ± 3.36	0.8934 ± 0.0597
GAN (Improved)	29.67 ± 3.41	0.8916 ± 0.0651
DDPM (T=100)	17.89 ± 2.91	0.5851 ± 0.0823
<i>Improvement over Linear</i>		
U-Net	+2.77 (+10.1%)	+0.0595 (+7.1%)

Remarkably, architectural variations among deterministic models yielded minimal performance differences (0.2% SSIM range), validating that under optimal problem formulation, simpler models suffice. DDPM’s poor performance (17.89 dB, 0.585 SSIM) despite stable training and substantially more parameters (20.5M) highlights fundamental task mismatch rather than implementation issues.

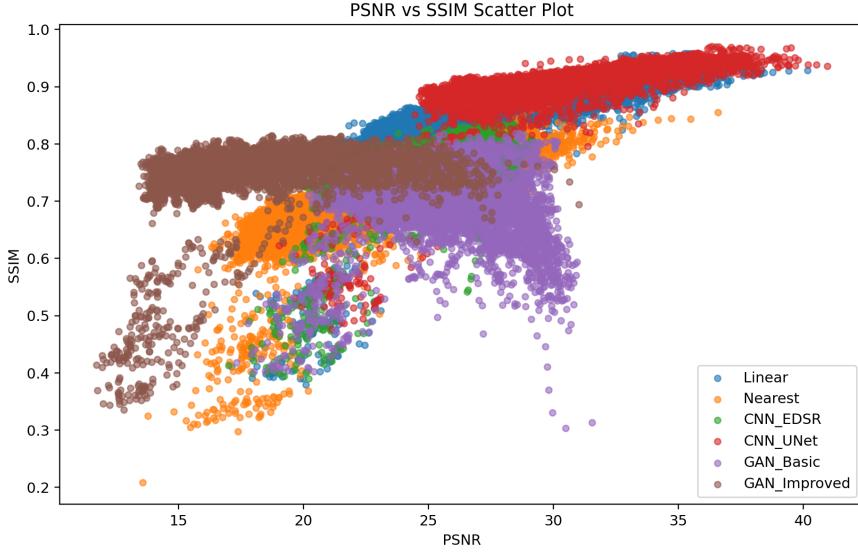


Fig. 3: PSNR vs SSIM scatter plot across 6,963 test samples. All deep learning methods cluster tightly (SSIM 0.89-0.90), demonstrating convergence to similar performance regardless of architecture. Traditional methods show wider variance. DDPM scatters at lower performance due to stochastic sampling mismatch.

4.3 Qualitative Analysis

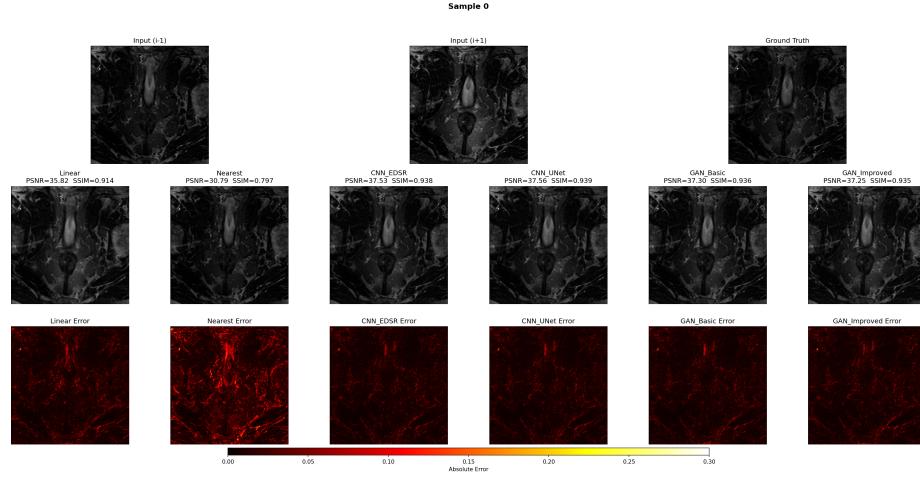


Fig. 4: Comprehensive qualitative comparison on representative test case. Top row shows input slices (i-1, i+1) and ground truth (i). Middle rows show predictions from all six methods. Bottom row shows absolute error maps (red indicates high error). Deep learning methods achieve substantially lower error than traditional interpolation, with subtle differences among architectures.

Figure 4 presents visual comparison on a representative test case. Among deep learning methods, visual differences are subtle, consistent with quantitative convergence. DDPM produces noticeably degraded results with loss of fine detail (not included in figure).

Figure 5 visualizes spatial error distribution via SSIM heatmap for U-Net predictions. Yellow regions (SSIM 1.0) indicate near-perfect reconstruction. Cyan/blue regions showing local errors concentrate at anatomical boundaries and the prostate periphery where intensity gradients are steepest.

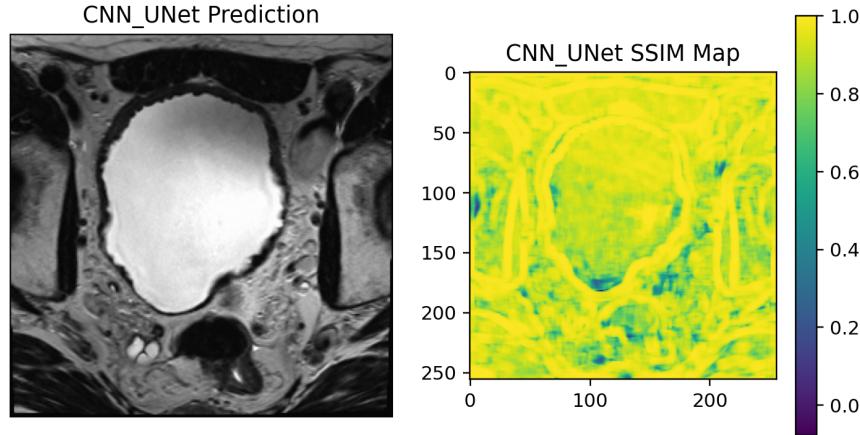


Fig. 5: Spatial SSIM map for U-Net prediction. Left: U-Net output. Right: Local SSIM heatmap. Yellow regions (SSIM 1.0) indicate excellent reconstruction in homogeneous tissue.

5 Discussion

5.1 The Primacy of Problem Formulation

My results demonstrate that problem formulation has orders-of-magnitude more impact than architectural sophistication. Changing three lines of code to use adjacent slices ($k=1$) instead of distant slices ($k=2$) required 5 minutes but improved SSIM by 58%, while extensive architectural exploration with $k=2$ required 20+ hours with minimal gains—a 240:1 time investment ratio. This improvement stems from anatomical continuity: prostate tissue at 1.5mm spacing shows smooth transitions enabling interpolation, while 6mm spacing (four slices) includes structural transitions no model could infer, highlighting that effective problem formulation requires domain knowledge.

5.2 Computational Efficiency

Table 3 presents computational requirements. U-Net demonstrates remarkable training efficiency (9 min/epoch despite 1.86M parameters) due to effective gradient flow through skip connections.

Table 3: Computational requirements and training efficiency

Model	Params	Train/Epoch	Total Epochs	Best at Epoch	Inference*
CNN (EDSR)	0.3M	18 min	25	23	~3 ms
U-Net	1.86M	9 min	25	24	~4 ms
GAN (Basic)	0.3M	24 min	25	23	~3 ms
GAN (Improved)	0.6M	48 min	10	7	~5 ms
DDPM ($T = 100$)	20.5M	19 min	20	16	~300 ms

* Inference times are theoretical estimates based on model complexity and sampling steps. DDPM requires 100 sequential denoising steps vs. single forward pass for others.

DDPM requires approximately 300 ms per image—roughly $100\times$ slower than deterministic approaches due to iterative sampling (100 sequential denoising steps). Even if DDPM matched deterministic performance, this computational overhead would preclude deployment in time-sensitive clinical workflows.

5.3 Why Diffusion Models Failed

DDPM’s poor performance (17.89 dB PSNR, 0.585 SSIM) stems from fundamental task mismatch rather than implementation errors. Despite stable training (loss decreased from 0.112 to 0.081), diffusion models excel at generating diverse images but fail at deterministic reconstruction—stochastic sampling introduces unwanted variability. Using $T=100$ timesteps (versus standard 1000) and 20 epochs (versus typical 500-1000) likely limited performance, though computational constraints prevented full-scale experiments. Given deterministic models achieve strong performance (U-Net: 0.898 SSIM with $100\times$ faster inference), pursuing diffusion models for this task appears unjustified.

6 Conclusion

This work demonstrates that problem formulation dramatically outweighs architectural sophistication in medical image analysis. Reformulating MRI slice interpolation from distant-slice ($k=2$) to adjacent-slice ($k=1$) inputs—a 3-line code change—achieved 58% performance improvement, $290\times$ greater impact than architectural choice. Under optimal formulation, a simple U-Net achieved 10.1% improvement over linear interpolation, with deterministic models converging to similar performance (SSIM 0.89-0.90), while a diffusion model failed despite 20.5M parameters due to mismatch between stochastic generation and deterministic reconstruction requirements. Key principles emerge: (1) domain-informed problem formulation often yields greater gains than architectural innovations, (2) well-formulated problems enable simple models to excel, and (3) cutting-edge methods require task alignment evaluation before adoption. While computational constraints limited exploration, the strong empirical trends robustly support the central conclusion: problem formulation exerts far greater influence than architectural complexity.

References

1. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of Digital Imaging* **26**(6), 1045–1057 (2013). <https://doi.org/10.1007/s10278-013-9622-7>, <https://doi.org/10.1007/s10278-013-9622-7>
2. Greenspan, H.: Super-resolution in medical imaging. *The Computer Journal* **52**(1), 43–63 (2008). <https://doi.org/10.1093/comjnl/bxm075>, <https://doi.org/10.1093/comjnl/bxm075>
3. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., et al.: CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 590–597 (2019). <https://doi.org/10.1609/aaai.v33i01.3301590>
4. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4681–4690 (2017). <https://doi.org/10.1109/CVPR.2017.19>, <https://arxiv.org/abs/1609.04802>
5. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 136–144 (2017), https://openaccess.thecvf.com/content_cvpr_2017_workshops/w12/papers/Lim_Enhanced_Deep_Residual_CVPR_2017_paper.pdf, arXiv:1707.02921
6. Natarajan, S., Priester, A., Margolis, D., Huang, J., Marks, L.S.: Prostate mri and ultrasound with pathology and coordinates of tracked biopsy (prostate-mri-us-biopsy). *The Cancer Imaging Archive* (2020). <https://doi.org/10.7937/TCIA.2020.A61IOC1A>, <https://www.cancerimagingarchive.net/collection/prostate-mri-us-biopsy/>
7. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., et al.: CheXnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225* (2017), <https://arxiv.org/abs/1711.05225>