Multi-Image Super-Resolution for Medical Slice Interpolation

Deivanai Thiyagarajan[1]

University of Florida, Gainesville, FL, USA
Master's in Applied Data Science
dthiyagarajan@ufl.edu

Abstract. This study compares deterministic and generative approaches for multi-image super-resolution to synthesize missing prostate MRI slices in a dataset with 3 mm and 6 mm inter-slice spacing. Evaluated models include a baseline Deep CNN, several UNet variants (MSE, combined perceptual/structural loss, progressive UNet, and GAN-based UNet), and two Fast-DDPM configurations. UNet-based methods achieve the highest reconstruction fidelity, while diffusion models lag despite producing anatomically coherent outputs. Training and testing across both gap sizes show that leveraging multi-slice context and richer loss functions improves reconstruction quality, whereas current fast diffusion setups struggle to match deterministic UNet performance.

Keywords: Super-resolution · UNet · GAN · Fast-DDPM

1 Introduction

High-resolution volumetric imaging is crucial in prostate MRI for accurate lesion localization, treatment planning, and biopsy guidance. Clinical protocols often acquire anisotropic volumes with large inter-slice spacing (e.g., 3 mm or 6 mm) to reduce scan time, which lowers through-plane resolution and can obscure small lesions. Synthesizing intermediate slices (slice interpolation / multi-image super-resolution) can enhance apparent z-resolution without extra scanning, but requires perceptual fidelity and preservation of diagnostically relevant structures.

This study investigates multi-image super-resolution for prostate MRI biopsy data with 3 mm and 6 mm gaps, conditioning on adjacent slices to leverage z-axis context. We implement and compare deterministic UNet-based approaches (baseline DeepCNN [1], UNet with MSE, UNet with combined perceptual/SSIM loss, progressive UNet [6], and UNet with adversarial training) against generative diffusion models [2] (Fast-DDPM variants). Models are evaluated quantitatively (PSNR, SSIM) and qualitatively for anatomical detail preservation.

Key challenges include handling variable slice spacing, maintaining clinical fidelity beyond simple image metrics, and balancing model complexity with inference speed. Experiments address these via spacing-aware training, loss functions combining pixelwise and structural similarity, and a fast diffusion scheduler [4]. Major contributions include:

- Comparative evaluation of UNet variants and fast diffusion models [4] for multi-image slice synthesis on prostate MRI with 3 mm and 6 mm spacing.
- Detailed implementation and training strategies (loss design, progressive staging, adversarial training, fast DDPM [4] scheduler).
- Quantitative (PSNR, SSIM) and qualitative analyses showing UNet variants achieve high reconstruction fidelity.

2 Related Work

- Image super-resolution and slice interpolation - Deep learning and video-interpolation methods, which map low- to high-resolution images and exploit temporal continuity [3], have been adapted to through-plane MRI, using neighboring slices as multi-image context.
- Adversarial and perceptual losses - GANs and perceptual losses (using features from pre-trained networks) enhance high-frequency detail and perceptual realism in super-resolution (e.g., SRGAN). In medical imaging, adversarial training can improve visual quality but may introduce hallucinated features, so combining pixelwise (MSE) and structural/perceptual (SSIM, VGG-based) losses is recommended.
- Diffusion and score-based generative models - Diffusion models [2] (DDPM/score-based) achieve state-of-the-art generative quality and are being applied to medical image synthesis. Their iterative denoising yields diverse, high-fidelity outputs, but inference is costly, so fast/reduced-step schedulers and conditional diffusion architectures are used to balance speed and quality.

3 Dataset

Our study uses a clinical Prostate MRI–Ultrasound (US) Fusion Biopsy dataset, focusing specifically on T2-weighted (T2w) MRI volumes. T2w MRI provides high soft-tissue contrast and is the preferred modality for prostate anatomical assessment, making it ideal for slice-interpolation and multi-image super-resolution research. The dataset includes a total of 1,151 patients, of which approximately 840 have T2w MRI scans available.

These scans have anisotropic resolution, with fine in-plane spacing (0.664×0.664 mm) but coarse through-plane (Z-axis) spacing ranging from 1.5 mm to 6 mm. This gap between slices leads to missing anatomical details, which can negatively affect visualization and diagnostic performance. Each image volume follows a consistent acquisition geometry with an identity direction matrix.

To create training data for slice interpolation, we extract paired triplets from each MRI volume. For short-range interpolation, slices i and $i+2$ are used to predict the intermediate slice $i+1$. To simulate more challenging 6-mm spacing, we also extract long-range triplets where slices i and $i+4$ are used to predict the middle slice $i+2$. Using both configurations helps the model learn anatomical continuity across different slice gaps.

Before training, all volumes are normalized using per-volume z-score normalization to make intensity distributions consistent across subjects. Slices are spatially aligned and resized to enforce a uniform dimension across the dataset. The augmentation includes random horizontal flips and small rotations between -5° and $+5^\circ$, chosen to preserve anatomical structure while providing meaningful variability.

Model performance is evaluated using two standard image quality metrics: Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR), computed on the predicted slices.

4 Methods

To address MRI slice interpolation, we benchmarked diverse architectures including convolutional, multi-stage, adversarial, and diffusion-based methods. Our baselines consist of a residual DeepCNN [1] and an MSE-trained UNet that use adjacent slices to model anatomical continuity. We further examined a UNet with a combined loss (MSE + perceptual + SSIM) for improved texture and structural fidelity, and a UNet-GAN [5] to promote sharper, more realistic reconstructions. For large inter-slice gaps, we adopted a Progressive UNet pipeline [6], where successive UNets iteratively refine intermediate predictions. We also evaluated Fast-DDPM [4], a lightweight diffusion model with reduced linear or cosine timesteps for efficient middle-slice synthesis.

All models take two neighboring slices as input ($2 \times 256 \times 256$) to predict the center slice ($1 \times 256 \times 256$), with Fast-DDPM additionally conditioned on timestep-dependent noise. Training was standardized across methods using batch size 4, 20 epochs, and Adam optimization. Evaluation on separate 2-slice and 4-slice gap test sets reports average loss, SSIM, and PSNR using final model predictions.

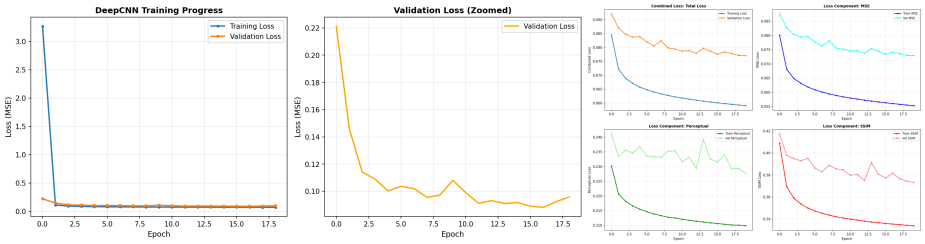

Fig. 1. DeepCNN (left) and UNet(right) Training Curve.

4.1 DeepCNN

The baseline model is a fully convolutional ResNet-style DeepCNN [1]. The encoder begins with a 7×7 convolution, batch-norm, ReLU and light pooling, fol-

lowed by four residual stages built from 3×3 conv residual blocks (configuration [2,2,2,2]) with channel expansion controlled by a base feature width (64). Residual connections and batch normalization are used throughout to stabilize training and preserve low-level detail; a 1×1 output convolution maps back to a single intensity channel so the model is fully convolutional and preserves spatial resolution.

Training optimizes a pixelwise mean squared error (MSE) between the predicted middle slice and the ground-truth slice. The MSE objective directly penalizes intensity differences and encourages minimization of average squared reconstruction error. During development the MSE validation loss was used for model selection. The training curve for the DeepCNN model is shown in Figure 1.

4.2 UNet

The UNet is a standard 2D encoder–decoder with double-conv blocks and batch normalization in each block, four downsampling pools, a bottleneck, and symmetric transpose-convolution upsampling with skip-connections. The input is the concatenated prior and posterior slices (2 channels), and the network outputs a single reconstructed middle slice (1 channel). The implementation uses `init_features` of 64 with feature multipliers $1, 2, 4, 8, 16$ at the bottleneck and preserves spatial resolution via a final 1×1 convolution, making the model fully convolutional for 256×256 in-plane inputs. Training optimizes a pixelwise mean squared error (MSE) between the prediction and ground truth.

The combined-loss UNet uses the same architecture described above but is trained with a multi-term objective that includes MSE, perceptual loss, and SSIM. The perceptual loss is an L_1 loss computed on VGG16 feature maps, after repeating the grayscale input to 3 channels and normalizing with ImageNet statistics. SSIM is implemented as a differentiable 1−SSIM Gaussian-window term. A conservative weighting was used: full weight for MSE and small weights for perceptual and SSIM losses with 0.1 each, with alternative balanced or aggressive configurations explored during development. . The training curve for the UNet model is shown in Figure 1.

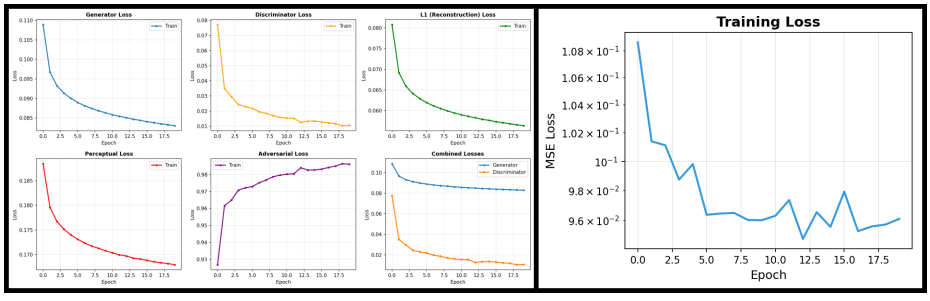

Fig. 2. UNet GAN (left) and Fast DDPM (right) Training Curve.

4.3 UNet - GAN

The generator is the same UNet architecture described above, taking the concatenated prior and posterior slices (2 channels) as input and producing a single reconstructed middle slice (1 channel). The discriminator follows a PatchGAN design with consecutive 4×4 convolutional blocks, LeakyReLU, and batch normalization, producing patch-wise real/fake scores to focus adversarial feedback on local texture and anatomical detail. Feature widths start from a base of 64 with standard $2\times$ channel growth in deeper stages, keeping the model fully convolutional for 256×256 inputs.

Training optimizes a composite generator loss that combines reconstruction loss, perceptual loss, and adversarial loss. In experiments, the reconstruction loss was given full weight, while perceptual and adversarial losses were given smaller weightings (10% and 1% of the reconstruction loss, respectively). The discriminator was trained using the corresponding LSGAN loss. The training curve for UNet-GAN [5] is shown in Figure 2.

4.4 Fast DDPM

The Fast-DDPM [4] model implements a conditional denoising diffusion approach for middle-slice reconstruction. The model uses a compact 2D UNet (double-conv blocks, time-MLP conditioning, base channels 64, time embedding dimension 256) and a FastNoiseScheduler that maps a 1000-step DDPM schedule to a smaller number of steps (e.g., T=10) with a non-uniform 40% early / 60% late split and sinusoidal timestep embeddings. The network is trained to predict noise in the latent space, using triplets of previous, next, and middle slices as input, and generates samples via a deterministic DDIM-style reverse pass.

The best model is selected based on validation SSIM. Evaluation computes PSNR and SSIM for generated slices, including stratified test splits for different inter-slice distances (distance=2 and 4), and visualizes predictions alongside training curves. The implementation emphasizes fast inference by reducing the number of diffusion steps while retaining the standard DDPM noise-prediction objective. The training curve for Fast DDPM [4] is also shown in Figure 2.

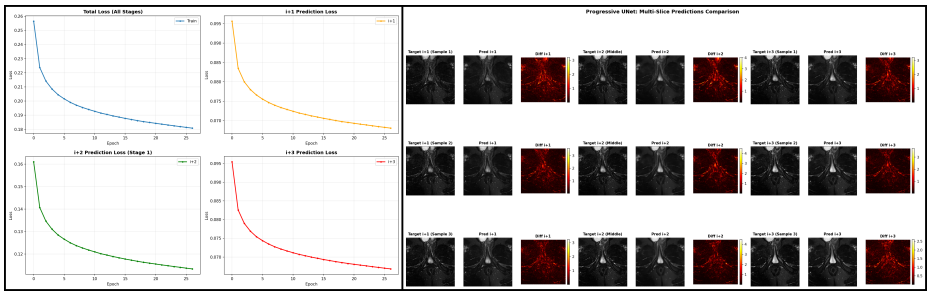

Fig. 3. Progressive UNet Training Curve and Outputs.

4.5 Progressive UNet

The Progressive UNet [6] is a three-stage cascaded UNet that explicitly models intermediate slices. Stage 1 predicts the central slice from the two far neighbors, while Stage 2 uses the Stage-1 output to predict the adjacent slices in parallel. Each stage is a standard 2D UNet (double 3×3 conv + BN + ReLU, four downsampling stages, bottleneck, transpose-conv upsampling with skip connections) with a final 1×1 convolution. The model takes slices shaped $(B, 5, H, W)$ and outputs three single-channel predictions for the intermediate slices. This design enforces consistency across neighboring slices and allows the central prediction to guide adjacent reconstructions.

Training uses a multi-scale mean squared error loss that combines reconstruction errors from all three predicted slices, with the central slice given the highest weight. Default loss weights were 0.5 for the adjacent slices and 1.0 for the central slice. The training curve for Progressive UNet [6] is also shown in Figure 3.

5 Results

Table 1. Quantitative comparison of all models on 3 mm and 6 mm slice gaps. Metrics reported as mean SSIM and PSNR (dB).

Model	SSIM(3 mm)	PSNR(3 mm)	SSIM(6 mm)	PSNR(6 mm)
DeepCNN	0.8217	26.30	0.5940	20.83
MSE UNet	0.8797	29.21	0.6530	21.91
Combined-Loss UNet	0.8804	29.21	0.6586	22.23
UNet-GAN [5]	0.8808	29.14	0.6574	22.08
Progressive UNet (avg)	0.7958	25.8	0.6645	22.44
Fast-DDPM [4] (T=10)	0.7590	24.14	0.4920	21.78

5.1 Quantitative Results

Across all models, performance on the 3 mm gap was consistently higher than on the more challenging 6 mm gap. As shown in Table 1, the MSE UNet, Combined-Loss UNet, and UNet-GAN [5] form the top-performing cluster for the 3 mm setting, reaching SSIM values near 0.88 and PSNR around 29 dB. DeepCNN trails behind with lower fidelity, while Fast-DDPM [4] achieves moderate accuracy but remains below the UNet baselines due to its limited diffusion timesteps.

For the 6 mm gap, reconstruction quality decreases for every model, reflecting the increased difficulty of predicting slices farther from the input. Notably, the Progressive UNet [6] achieves the highest SSIM and PSNR for 6 mm spacing (Table 1), outperforming all single-stage models. Although its improvements

are modest, the multi-stage progression and intermediate refinements help preserve structure more effectively, demonstrating a clear advantage in wider-gap prediction.

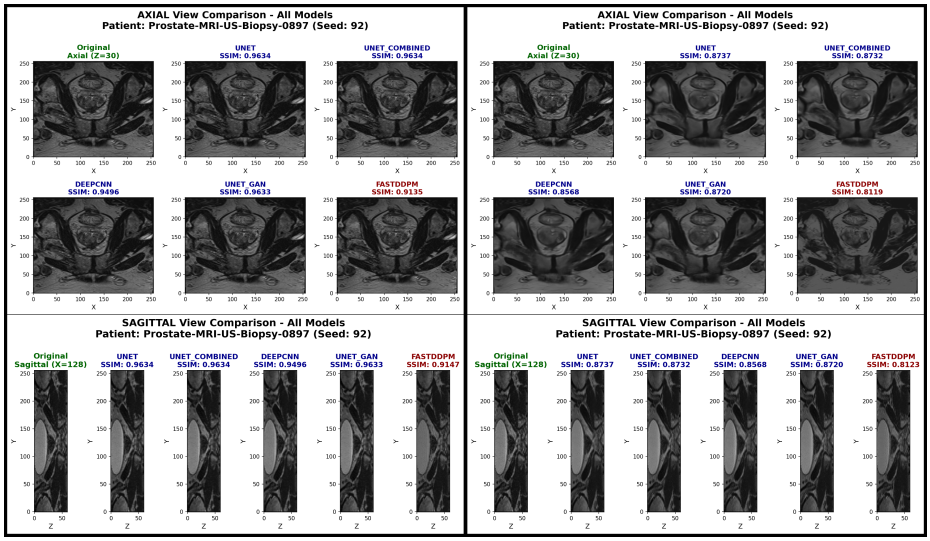

Fig. 4. Axial and Sagittal Views 3mm(left) and 6mm(right) spacing interpolation.

5.2 Qualitative Analysis

Figure 4 presents axial and sagittal reconstructions for a representative patient. The visual trends align well with the quantitative results: the MSE UNet, Combined-Loss UNet, and UNet-GAN [5] generate the cleanest and most accurate slices, with clear anatomical boundaries and minimal smoothing. Their outputs closely match the ground truth in both views.

DeepCNN produces noticeably softer and blurrier textures, while Fast-DDPM [4], although structurally consistent, shows reduced sharpness—especially in the 6 mm gap case, where details appear more smoothed out. Overall, the UNet-based models preserve structure far better than both the CNN and diffusion baselines.

Among the top-performing models, UNet-GAN [5] and Combined-Loss UNet provide slightly improved texture realism compared to the pure MSE UNet, particularly in low-contrast regions and sagittal slices.

Figure 5 presents a side-by-side comparison of single-triplet predictions from the three top-performing models (selected based on average SSIM and PSNR). Each row corresponds to a model (top: UNet, middle: UNet Combined, bottom: UNet GAN [5]), while the columns show (left -> right) the preceding slice (PRE,

slice 26), the following slice (POST, slice 28), the ground-truth middle slice (slice 27), and the model's predicted middle slice.

Visually, all three models recover major anatomical structures (prostate and surrounding tissue) accurately, with subtle differences in local contrast and edge sharpness. UNet Combined appears slightly closer in global intensity to the ground truth, while UNet GAN [5] preserves overall detail without a noticeable advantage, suggesting that any perceptual improvements [5] from the GAN are small or not captured visually. The uniformly consistent predictions indicate that remaining differences are mainly low-amplitude smoothing rather than significant structural inaccuracies.

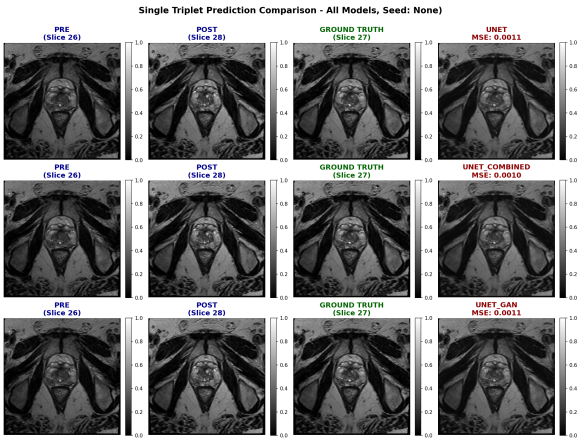

Fig. 5. Comparison of predicted middle slices from the top three models.

6 Discussion and Conclusion

Our study shows that UNet-based models (MSE, Combined-Loss, and GAN variants) consistently outperform Fast-DDPM [4] diffusion models in prostate MRI slice interpolation. The UNet with Combined Loss achieved the highest reconstruction quality (SSIM 0.7714, PSNR 25.78 dB averaged across both spacing), while Fast-DDPM [4] lagged significantly (SSIM 0.6371, PSNR 22.85 dB averaged across both spacing). These results indicate that direct regression approaches are better suited than generative models for accurately reconstructing missing slices in clinical MRI volumes.

Adversarial training via UNet-GAN [5] provided only marginal gains over standard UNet, and incorporating combined loss functions (MSE + perceptual + SSIM) consistently improved anatomical fidelity and texture preservation. The quantitative and qualitative analyses highlight that multi-slice context and care-

fully designed loss functions are critical for high-quality slice synthesis, especially for larger inter-slice gaps.

From a computational perspective, Fast-DDPM [4] achieves faster inference due to reduced denoising steps but at the cost of lower reconstruction quality. UNet variants, although slower, provide more accurate and clinically useful outputs. In summary, deterministic UNet architectures with combined loss functions offer the best trade-off between accuracy and reliability for medical image interpolation, while future work could explore hybrid approaches to combine speed and perceptual quality.

7 Acknowledgement

The authors gratefully acknowledge the support and guidance provided by the course instructors throughout this project. We also thank the University of Florida for providing computational resources and storage infrastructure, which were essential for model training and experimentation. The implementation of all models and training pipelines used in this research is publicly available at: https://github.com/DeivanaiThiyagarajan/Multi-Image-Super-Resolution-for-Medical-Images/tree/main

References

1. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(2), 295–307 (2015)
2. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851 (2020)
3. Sood, R.R., Shao, W., Kunder, C., Teslovich, N.C., Wang, J.B., Soerensen, S.J.C., Madhuripan, N., Jawahar, A., Brooks, J.D., Ghanouni, P., et al.: 3D registration of pre-surgical prostate MRI and histopathology images via super-resolution volume reconstruction. Medical Image Analysis **69**, 101957 (2021), https://pubmed.ncbi.nlm.nih.gov/33550008
4. Jiang, H., Imran, M., Zhang, T., Zhou, Y., Liang, M., Gong, K., Shao, W.: Fast-DDPM: Fast Denoising Diffusion Probabilistic Models for Medical Image-to-Image Generation. arXiv preprint arXiv:2405.14802 (2024), https://arxiv.org/abs/2405.14802
5. Zhang, X., Li, H., Li, H., Ma, L., Wang, L., Huang, Y.: Multi-contrast super-resolution MRI through a progressive network. Magnetic Resonance Imaging **76**, 45–55 (2020), https://pubmed.ncbi.nlm.nih.gov/32086201
6. Liu, L., Chen, F., Xing, Y., Zhang, L., Wang, F., et al.: Progressive sub-band residual-learning network for MR image super-resolution. Magnetic Resonance Imaging **71**, 123–131 (2020), https://pubmed.ncbi.nlm.nih.gov/31603805
7. Goswami, S., Gupta, A., Paul, A.: LearnDiff: MRI image super-resolution using a diffusion model with learnable noise. (2025), https://www.sciencedirect.com/science/article/abs/pii/S0895611125001508