2026 JULY 19–22
VANCOUVER, BC
JOINT AAPM | COMP MEETING

Multi-Resolution Patch-Based Diffusion Priors for Short-Arc CBCT Reconstruction

Kwok Lung Fan[1], Yabo Fu[1], Licheng Kuo[1], Jiening Zhu[1], Jean Moran[1], Laura Cervino[1], Xiang Li[1], Tianfang Li[1]
[1]Memorial Sloan Kettering Cancer Center

ABSTRACT

Purpose: To develop a short-arc kV-imaging based intrafraction motion monitoring that enables reconstruction of 3D volumes of the prostate and nearby organs at risk (OARs) within seconds during treatment. We propose an iterative short-arc CBCT reconstruction framework regularized by a multi-resolution, patch-based diffusion prior to address the severe ill-posedness caused by limited-angle acquisition.
Methods: Two population-based diffusion priors were trained using image patches extracted from pelvic CBCTs of 260 patients: a 64 × 64 patch model for coarse reconstruction and a 32 × 32 patch model for refinement. Short-arc projection data spanning 90°, 60°, 45°, and 20° were simulated from planning CTs. Prostate-focused reconstruction was performed using diffusion posterior sampling with denoising diffusion implicit model (DDIM)-style updates and measurement-consistency steps within a prostate ROI defined by the PTV. Pixels outside the ROI were constrained using a strong prior derived from a setup full-arc CBCT, allowing the evaluation to focus on prostate and OAR depiction. A coarse-to-fine strategy was employed, in which the coarse model captured global anatomy and the refinement model recovered boundaries and fine details, while data fidelity was enforced through the measurement model. Reconstructions were compared with an iterative baseline using the same surrounding prior, with performance evaluated using the structural similarity index measure (SSIM).
Results: Diffusion-regularized reconstructions showed improved qualitative prostate conspicuity relative to the baseline across all angular spans. Reconstructions using the coarse model were completed in approximately 5 s per volume on a modern GPU. For the 90° case, SSIM increased from 0.876 to 0.901 with the diffusion-regularized reconstruction. Multi-resolution refinement improved edge sharpness at moderate angular coverage levels of 90°–45°. At severe angular sparsity, especially at 20°, reconstructions showed increased sensitivity to model-driven hallucination.
Conclusion: Diffusion-based posterior sampling is feasible for prostate-focused reconstruction from short-arc CBCT. A multi-resolution patch-based diffusion prior can enhance structural visibility under limited-angle conditions. These findings support further development toward markerless intrafraction monitoring and on-treatment assessment applications.

CONTACT

Tianfang Li, PhD, DABR
Memorial Sloan Kettering Cancer Center
1101 Hempstead Tpke, Uniondale, NY 11553
lit2@mskcc.org
631-623-4229

INTRODUCTION

Real-time target tracking during prostate volumetric modulated arc therapy (VMAT) is vital for ensuring precise dose delivery while sparing adjacent organs at risk (OARs). Although on-treatment kV imaging provides an ideal window for localization, clinical time constraints restrict these acquisitions to short-arc projections (<90°). Under these severe limited-angle conditions, reconstruction is highly ill-posed, meaning that infinitely many solutions can satisfy the measured data and prior information is needed. Generative diffusion models can address this by encoding the underlying anatomical data distribution and using the reverse-process score function to steer reconstruction toward a valid image manifold [1,4]. However, learning a full-field-of-view (FOV) pelvic manifold is difficult, as high structural variability requires much larger datasets to achieve convergence and reduce unphysical hallucinations. To address this, we propose a region-restricted population prior framework that focuses a pretrained U-ViT score network [3] exclusively on a localized region centered on the prostate. Restricting the network to this compact volume reduces the intrinsic dimensionality, enabling robust manifold characterization from smaller patient cohorts.

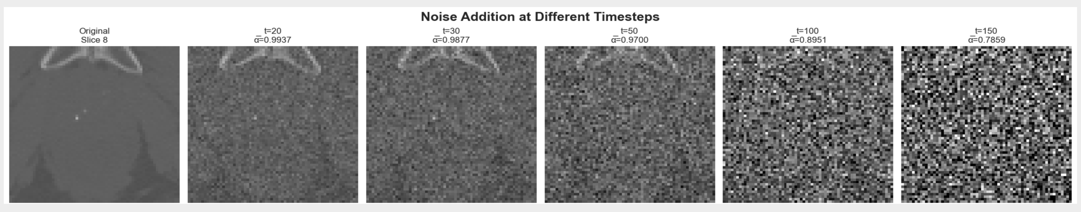

Figure 1. Diffusion process on prostate patches

METHODS AND MATERIALS

1. Data Preparation & Multi-Resolution Patches
•**Patient Cohort:** Pelvic CBCT datasets from 260 prostate cancer patients were utilized to capture realistic inter-fraction variations (e.g., rectal filling, bladder shifts) and modality-specific scatter artifacts.
•**Patch Extraction:** Localized patch grids were extracted around the prostate region of interest (ROI) to collapse the high structural dimensionality of full field-of-view abdominal structures.
•**Multi-Resolution Priors:** Two independent generative population models were trained via a U-ViT framework to facilitate a coarse-to-fine anatomical recovery strategy:
- **Coarse Model:** Trained on 64x64x patches to capture structural context.
- **Refinement Model:** Trained on 32x32 patches to recover sharp soft-tissue boundaries and fine details.

2. Short-Arc Acquisition Simulation
- **Projection Simulation:** Forward projection sinograms were simulated from planning CTs using the TIGRE[5] toolbox integrated with PyTorch-accelerated operators.
- **Primary Test Case:** Evaluation focused on a 90°
- limited-angle acquisition geometry as the primary test scenario, utilizing 120 uniform projection views.

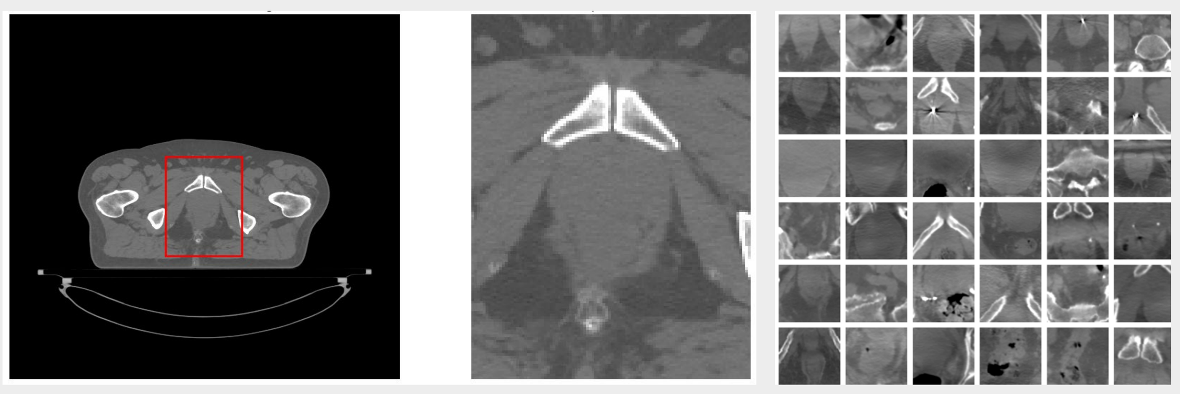
Figure 2. Example of training ROI and training patches (64x64)

METHODS AND MATERIALS (CONT)

Diffusion Posterior Sampling (DPS) Reconstruction
Score-Based Guidance: Reconstruction within the ROI was regularized by the patch diffusion priors, using the learned score function $(\nabla_{x_t} \log p(x_t))$ to guide updates through the underdetermined null space of A.
Conditional Update Loop: Denoising was performed via Denoising Diffusion Implicit Model (DDIM)-style updates[1,2] combined with image-space data-consistency gradient steps[6]:
- **Step 1:** Predict the clean patch estimate $\widehat{x_0}$ at current timestep t
- **Step 2:** Compute $\nabla_{\widehat{x_0}} p(y|\widehat{x_0})$ and apply an adaptive gradient update step λ_t to enforce projection data fidelity.
- **Step 3:** Fed the updated $\widehat{x_0}$ back into the stochastic DDIM sampling scheme to update the localized state to x_{t-1}

Peripheral Constraints: Pixels outside the prostate were constrained using a prior derived from a full-arc setup CBCT. This focused the generative capacity entirely on localized intrafraction target changes.
Coarse-to-Fine Multi-Resolution Refinement
Coarse Stage (Macro-Anatomy): Denoising is initialized at a high noise level of $t = 0.15 (\sigma \approx 733 \text{ HU})$. The 64x64 coarse model recovers global layout and low-frequency soft-tissue variations, driven by concurrent projection gradient updates.
Fine Stage (Boundary Refinement): A sequential quadrant update is executed using the 32x32 refinement model. A small noise envelope of $t = 0.01 (\sigma \approx 63 \text{ HU})$ is applied to preserve the macro-anatomy while recovering sharp soft-tissue boundaries and fine details.
Test Cases: We run our proposed method on 15 test cases and compare the results.

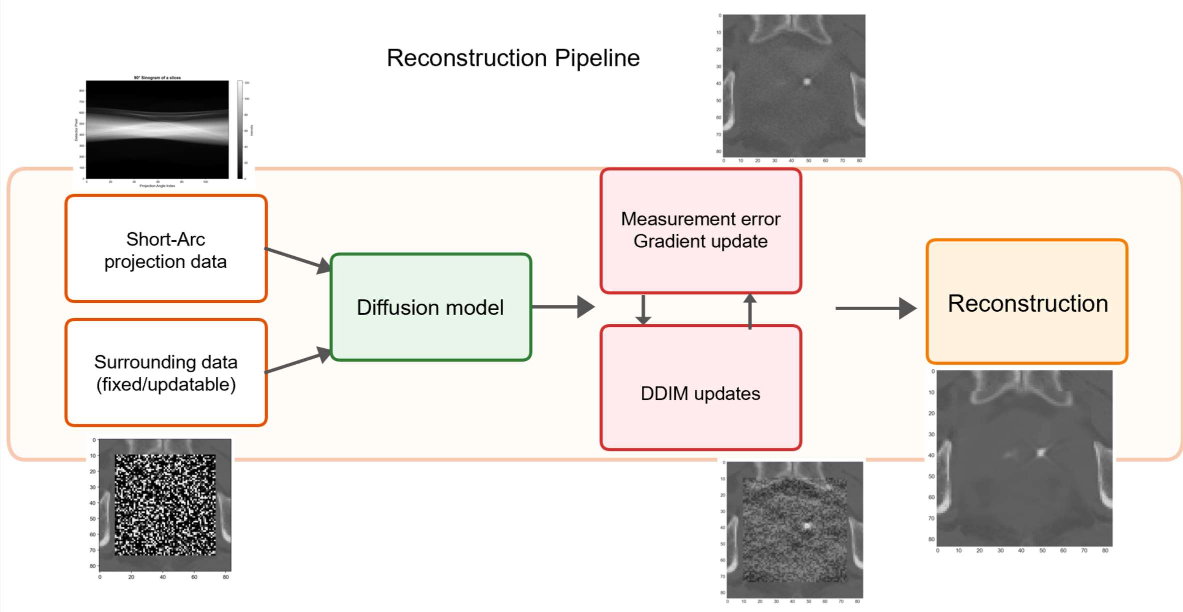

Figure 3. DPS reconstruction workflow

RESULTS

Method	PSNR (dB)	SSIM
Iterative (Baseline)	28.48 ± 2.14	0.7059 ± 0.0568
Diffusion (Coarse)	**32.28 ± 1.90**	**0.9040 ± 0.0357**
With Refinement	32.01 ± 1.94	0.8980 ± 0.0360

Table 1. Summary of result

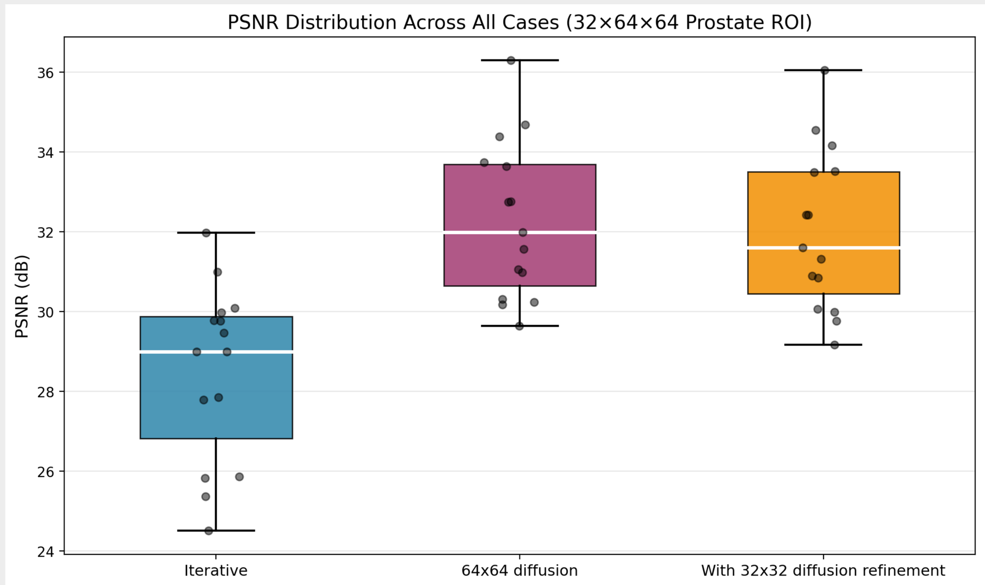

Figure 4. PSNR comparison between iterative reconstruction with total variation and the diffusion posterior sampling

RESULTS(CONT.)

Performance Breakdown: The patch-based generative diffusion frameworks demonstrated substantial performance gains over the standard iterative baseline across both pixel-level accuracy (PSNR) and structural fidelity (SSIM) within the 32x64x64 prostate ROI (**Table 1**, **Figure 4**).
The Metric vs. Perception Tradeoff: 64x64 coarse diffusion model achieved the highest global quantitative metrics, yielding a mean PSNR of 32.28 dB and an SSIM of 0.9040 (**Table 1**).
Complementary Structural Refinement: The addition of the 32x32 multi-resolution refinement stage resulted in a minor, statistically expected shift in global pixel-wise scores. However, the refinement stage improves the overall edge sharpness and boundary definition, recovering alternative soft-tissue details where global models tend to oversmooth (**Table 1**, **Figure 5**).
Model Hallucinations: Both models exhibited structural hallucinations. This likely stems from the population prior's broad anatomical solution space, which a 90° data constraint cannot completely bind.

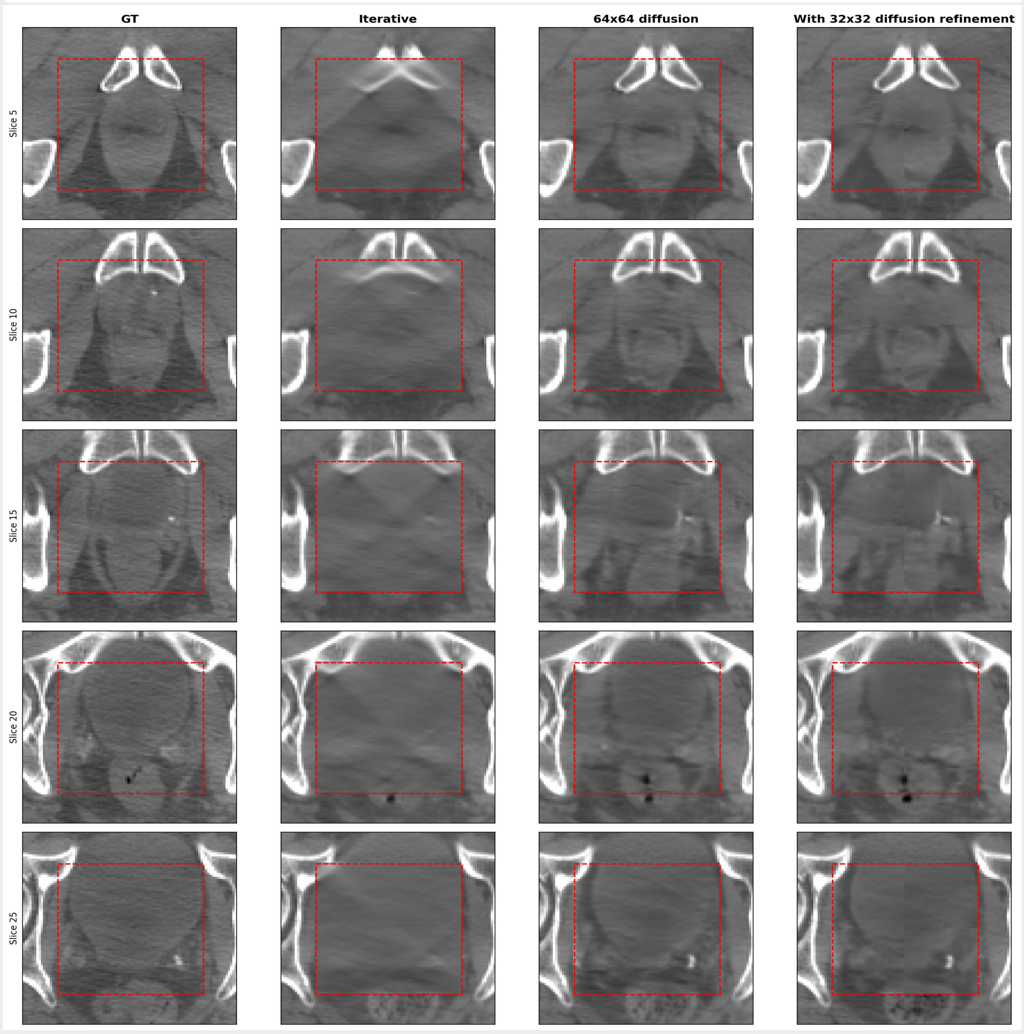

Figure 5. Reconstruction result visualization

CONCLUSIONS

Feasibility of Short-Arc CBCT This work demonstrates that a region-restricted, multi-resolution diffusion framework can successfully reconstruct target volumes from a limited-angle acquisition, outperforming traditional iterative techniques.
Task-Specific Tradeoffs While the coarse model provides the highest global pixel-wise metrics, the refinement stage improves localized edge sharpness and boundary definition along critical soft-tissue interfaces.
Mitigating Hallucinations Both models exhibit minor, expected hallucinations due to the broad anatomical solution space of a population-based prior under sparse angular constraints. Future work will focus on constraining result by incorporating patient-specific models to tighter restrict the solution space.

REFERENCES

1. Song, J., Meng, C., & Ermon, S. (2021). Denoising Diffusion Implicit Models. International Conference on Learning Representations. International Conference on Learning Representations.
2. Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., & Ye, J. C. (2023). Diffusion Posterior Sampling for General Noisy Inverse Problems. The Eleventh International Conference on Learning Representations . The Eleventh International Conference on Learning Representations .
3. Song, Y., Shen, L., Xing, L., & Ermon, S. (2021). Solving Inverse Problems in Medical Imaging with Score-Based Generative Models. NeurIPS 2021 Workshop on Deep Learning and Inverse Problems.
4. Bao, F., Li, C., Cao, Y., & Zhu, J. (2022). All are Worth Words: a ViT Backbone for Score-based Diffusion Models. NeurIPS 2022 Workshop on Score-Based Methods. NeurIPS 2022 Workshop on Score-Based Methods.
5. Biguri, A., Dosanjh, M., Hancock, S., & Soleimani, M. (2016). TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction. Biomedical Physics & Engineering Express, 2(5), 55010.
6. Jiang, X., Li, S., Teng, P., Gang, G., & Stayman, J. W. (2024). Strategies for CT reconstruction using diffusion posterior sampling with a nonlinear model. ArXiv, arXiv-2407.