A Conditional Wavelet Diffusion Framework with Depthwise Convolutions for 3D Medical Image Synthesis

Farshad Safavi, PhD[1]; Aman Sangal, PhD[1]; Hualiang Zhong, PhD[2] and Lei Ren, PhD[1]
[1]University of Maryland School of Medicine, Baltimore, MD, [2]Medical College of Wisconsin, Milwaukee

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

ABSTRACT

High-resolution 3D medical image synthesis is important for MRI modality completion, digital phantom generation, and radiotherapy applications.

This study proposes a conditional Wavelet Diffusion framework with depthwise separable convolutions for efficient and anatomically consistent MRI and CT synthesis. MRI translation and CT image generation were modeled as conditional image to image translation tasks using wavelet-domain diffusion with an efficient 3D denoising network using depthwise and pointwise convolutions.

The proposed framework achieved accurate MRI synthesis (SSIM: 0.912–0.952) and realistic CT generation (SSIM = 0.864, PSNR = 24.44 dB), demonstrating scalable diffusion-based synthesis for multimodal imaging and digital phantom applications.

CONTACT

Farshad Safavi
University of Maryland, Baltimore (UMB)
fsafavi1@umbc.edu

INTRODUCTION

Three-dimensional (3D) medical image synthesis has important applications in MRI modality completion, digital phantom generation, and radiotherapy planning. However, applying diffusion models to high-resolution volumetric images remains challenging due to the significant GPU memory and computational demands of conventional 3D architectures. To overcome these limitations, we propose a **Conditional Wavelet Diffusion Framework with Depthwise Convolutions** for efficient and anatomically consistent 3D medical image synthesis. The framework improves efficiency through two complementary strategies. First, wavelet-domain diffusion uses 3D Haar wavelet decomposition to reduce spatial complexity while preserving anatomical structures and high-frequency texture information (Friedrich et al., 2024):

$$y(H \times W \times D) \rightarrow DWT(y) = 8 \times \frac{H}{2} \times \frac{W}{2} \times \frac{D}{2}$$

Second, the denoising network is optimized using depthwise separable convolutions, which separate spatial feature extraction from channel mixing (Chollet, 2017):

$$3D\ Conv \ \rightarrow (Depthwise + Pointwise)\ Convs$$

This combined wavelet-domain and depthwise convolution strategy enables scalable high-resolution 3D diffusion modeling while maintaining anatomical fidelity for MRI synthesis and CT image generation.

METHODS AND MATERIALS

The proposed framework formulated 3D medical image synthesis as a paired conditional image to image translation problem for MRI cross-modality synthesis and CT texture generation (Figures 1 and 2). Input volumes were transformed into the wavelet domain using a 3D Haar Discrete Wavelet Transform (DWT), allowing diffusion to operate on compact multi-scale representations while preserving anatomical structures and texture details.

Experiments were performed on two datasets: BraTS 2024 MRI (1,350 training and 188 validation subjects with T1, T1ce, T2, and FLAIR modalities) and 317 clinical CT volumes for realistic phantom texture synthesis. All volumes were resized to $192^3$and normalized before training. The model was trained using a 1000-step diffusion process. Performance was evaluated using SSIM, PSNR, MAE, and qualitative multi-planar anatomical assessment.

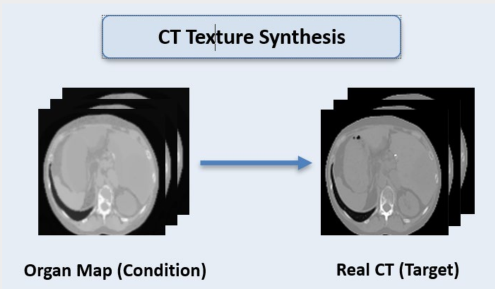


Figure 1. CT Texture Synthesis

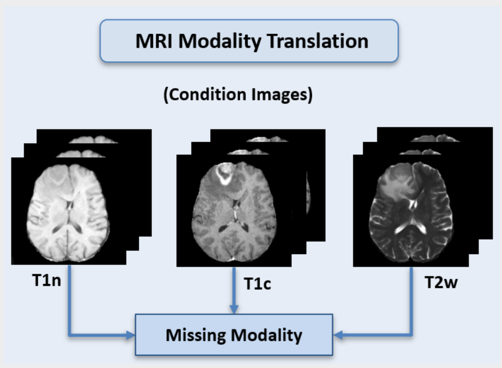


Figure 2. MRI Cross-Modality Synthesis

RESULTS

This new method generated anatomically realistic and structurally consistent 3D images across MRI modality completion and CT synthesis tasks (Figure3, Figure4).

MRI Modality Completion: The framework reconstructed missing MRI contrasts while preserving anatomical structures and modality-specific features. Quantitative results in Table 1 demonstrate high structural fidelity, while Figure 4 shows consistent 3D anatomical recovery across synthesized volumes.

CT Generation: The model generated realistic CT volumes from anatomy-preserving organ maps while maintaining tissue textures and anatomical boundaries. Quantitative results (Table 2) and qualitative examples (Figure 3) demonstrate strong agreement with reference CT images

Computational Efficiency: The proposed depthwise separable 3D network reduced computational complexity while maintaining synthesis performance:
- Parameters reduced by **up to 72.8%**
- Computational cost **reduced by up to 2.6×**

Table 1. Quantitative Evaluation of MRI Modality Completion

Target	SSIM	PSNR	MAE
T2-FLAIR	0.912	23.14	0.024
T2W	0.943	25.60	0.017
T1N	0.952	26.82	0.016

Table 2. Quantitative Evaluation of CT Texture Synthesis

Plane	SSIM	PSNR (dB)
Axial	0.855	23.20
Coronal	0.870	24.82
Sagittal	0.868	25.31
Mean	0.864	24.44

Figure 3. Qualitative Results of CT Texture Synthesis

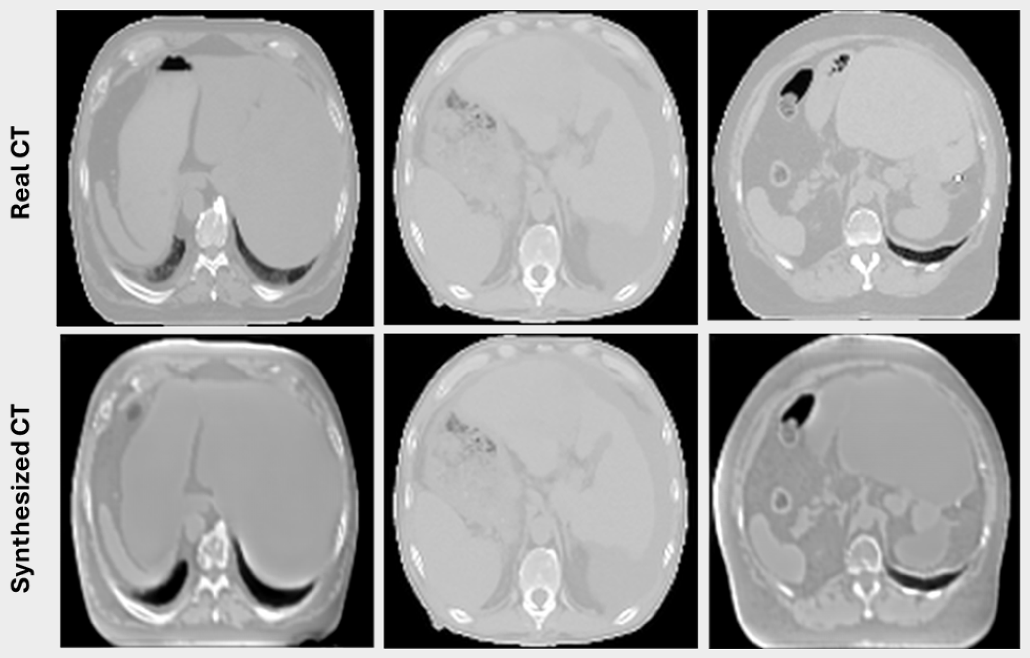
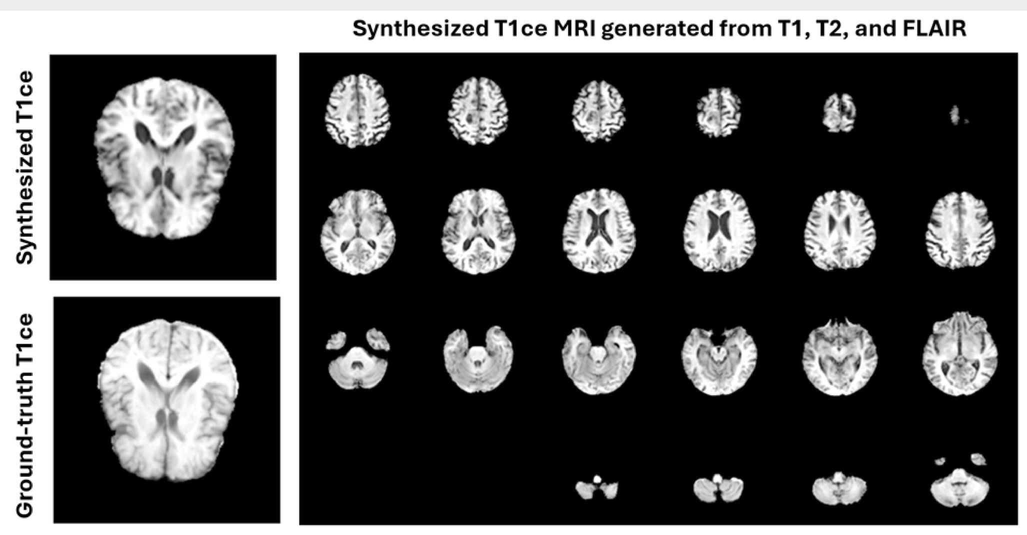


Figure 4. Qualitative Results of MRI Modality Completion



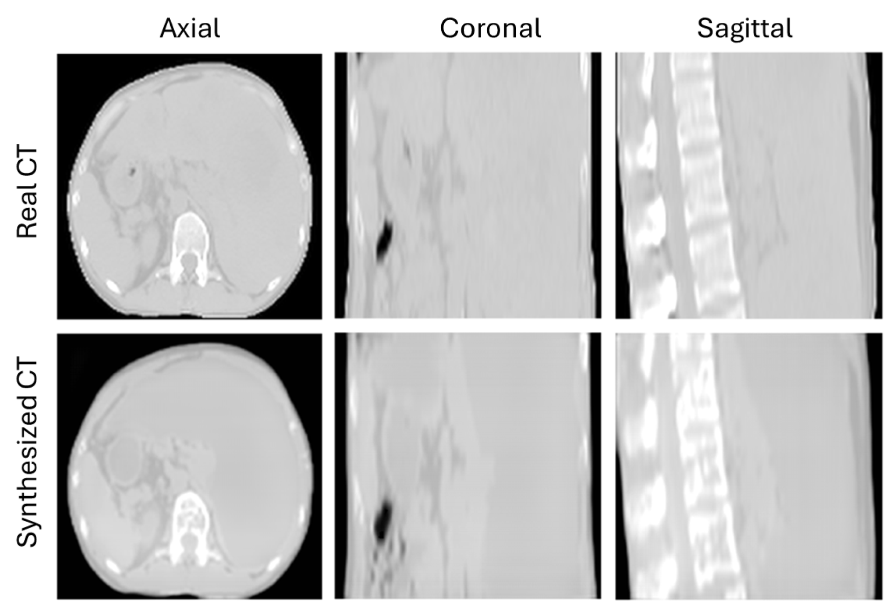
DISCUSSION

The proposed method generated anatomically consistent 3D volumes for MRI modality completion and CT texture synthesis. Multi-planar evaluation across axial, coronal, and sagittal views (Figure 5) demonstrates preservation of anatomical structures, spatial consistency, and fine texture details.

The improved efficiency is achieved by applying diffusion in the wavelet domain, where 3D Haar decomposition reduces spatial complexity while preserving anatomical information through multi-scale representations. Replacing conventional 3D convolutions with depthwise separable convolutions further reduces memory usage, parameters, and computational cost while maintaining image quality, enabling scalable 3D medical image synthesis.

Future work will extend this framework to additional MRI and CT synthesis tasks, including multimodal image generation and patient-specific digital phantom development.

Figure 5. Representative CT Texture Synthesis Results



CONCLUSIONS

This work presents a Conditional Wavelet Diffusion Framework with Depthwise Convolutions for efficient high-resolution 3D medical image synthesis. Experiments on BraTS MRI and clinical CT datasets demonstrated accurate MRI modality completion and realistic CT texture generation while preserving anatomical structures and volumetric consistency.

The framework achieved strong quantitative performance with high SSIM and PSNR values across synthesis tasks. The depthwise-separable 3D denoising network reduced model complexity by up to 72.8% and computational cost by 2.6× while maintaining synthesis quality. These results demonstrate the potential of the proposed framework for scalable intra and cross-modality synthesis, digital phantom generation, and radiotherapy planning applications.

REFERENCES

1. Friedrich et al., cWDM: Conditional Wavelet Diffusion Models for Cross-Modality 3D Medical Image Synthesis, arXiv:2411.17203, 2024.
2. Xception: Deep Learning with Depthwise Separable Convolutions, CVPR, 2017.