

Time-frequency Analysis in Generative AI:
From Wavelet-Based CycleGAN to DDWM
Preserving Structure in Medical Image Synthesis



R14942056 張允妍

The challenge in Medical Imaging

- High radiation dose = High quality image = CT
- Low radiation dose = Noisy image = low-dose CT / CBCT

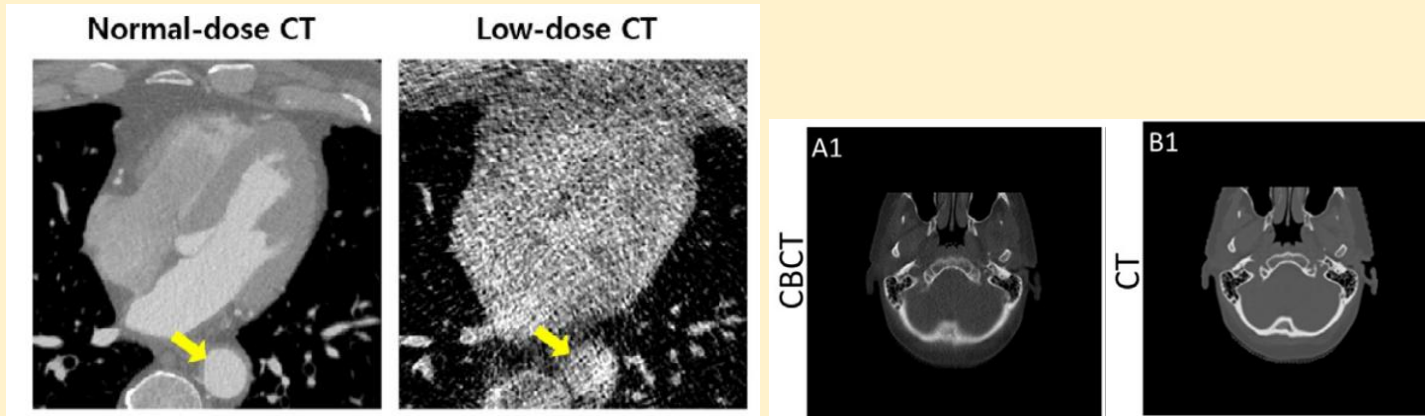


Fig. 1. (a) Normal-dose vs Low-dose CT (b) CBCT vs CT

● Goal

- to convert low-quality image (CBCT / Low-dose) into high-quality image (CT)

● Problem

- Therefore, generative models, e.g., GANs, VAEs, Diffusion Model, often suffer from “hallucinations”
- They might generate fake or misleading details, e.g., creating fake tumors or removing real blood vessels, cause they don't understand the “structure”.

● The solution: Time-Frequency Guidance

- Adding time-frequency into the pipeline to have a better guidance to the generating process
- Cause images are composed by high frequency(noise, edges, textures) and low frequency(anatomical structure, domain-invariant information)
- Using Wavelet Transform to separate the Structure(low freq.) from the Noise(high freq.)

CycleGAN denoising of extreme low-dose CT using wavelet-assisted noise disentanglement

- Task: Denoising “Extreme low-dose” CT image (only 4% dose)
- Model: CycleGAN (unsupervised)
- Problem: At 4% dose, cycleGAN algorithm often introduces artificial features
- Proposed architecture: cycleGAN with “wavelet-assisted noise disentanglement (WAND)”
 - Since low-dose CT noises are usually high frequency components, the proposed method first separates the low- and high- frequency components of images using wavelet transform

Wavelet-Assisted Noise Disentanglement(WAND)

● Method

1. Apply Discrete Wavelet Transform
2. Extract High-frequency subbands
3. Train the GAN only on these high-frequency maps to learn “what noise looks like” and remove it
4. add the cleaned high-frequency parts back to the original low-frequency parts

Table 2

Comparison of PSNR and SSIM for wavelet transform levels.

Level	2	4	6	8
PSNR	25.83	28.46	29.80	29.08
SSIM	0.4759	0.5635	0.6126	0.5965

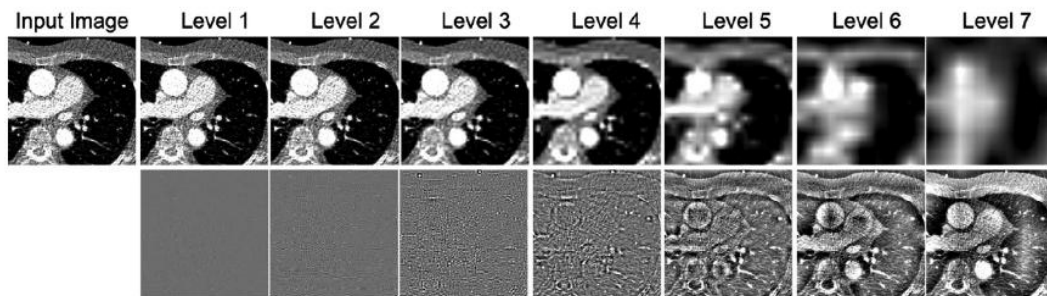


Fig. 3. Noise disentanglement by wavelet transform: the image in the leftmost column is a low-dose CT image. The images in the top row are the low-frequency wavelet images, and those in the bottom row are the high-frequency ones. From the second left column to the rightmost column, the level of the wavelet decomposition increases from 1 to 7. The image window is $(-700, 700)$ [HU].

Process for wavelet-assisted noise disentanglement

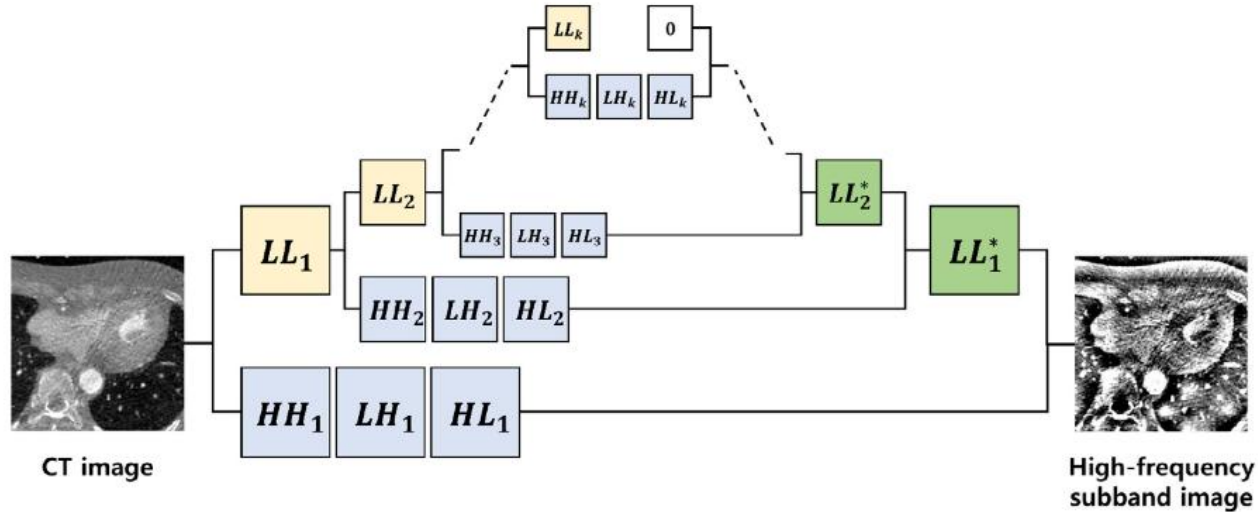


Fig. 4. The process for wavelet-assisted noise disentanglement by extracting high-frequency subband images from a CT image using k -level wavelet transform and inverse wavelet transform.

Training framework of the wavelet subband cycleGAN

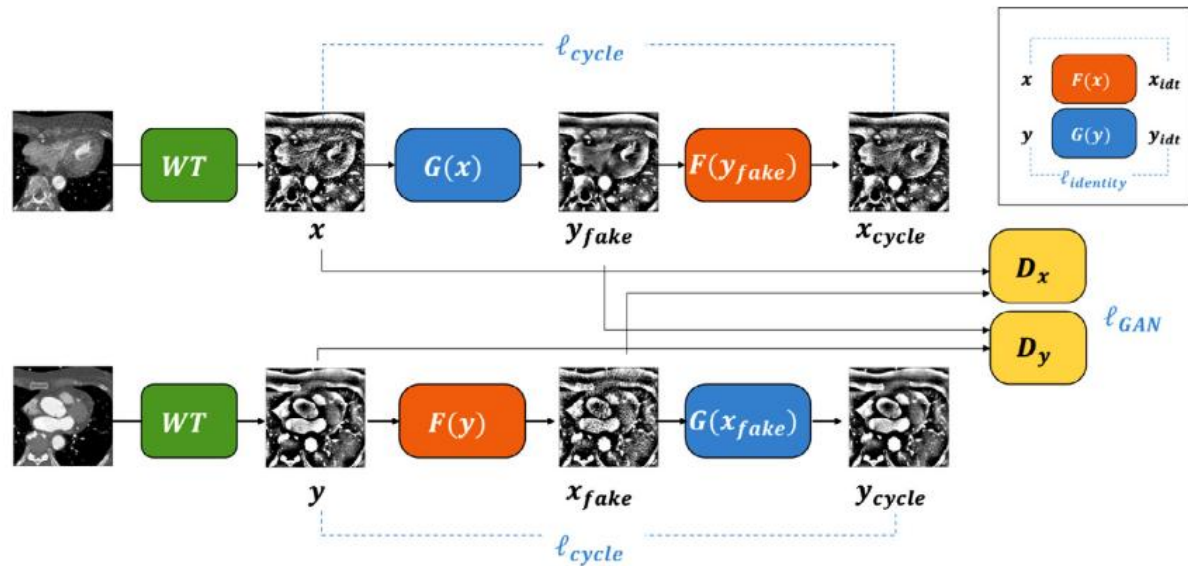


Fig. 5. Training framework of the wavelet subband cycleGAN. Firstly, high-frequency sub-band images are obtained from the CT images. There are two generators (F , G) and two discriminators (D_x , D_y). All of the networks are trained using the high-frequency subband images.

Inference step of wavelet based cycleGAN

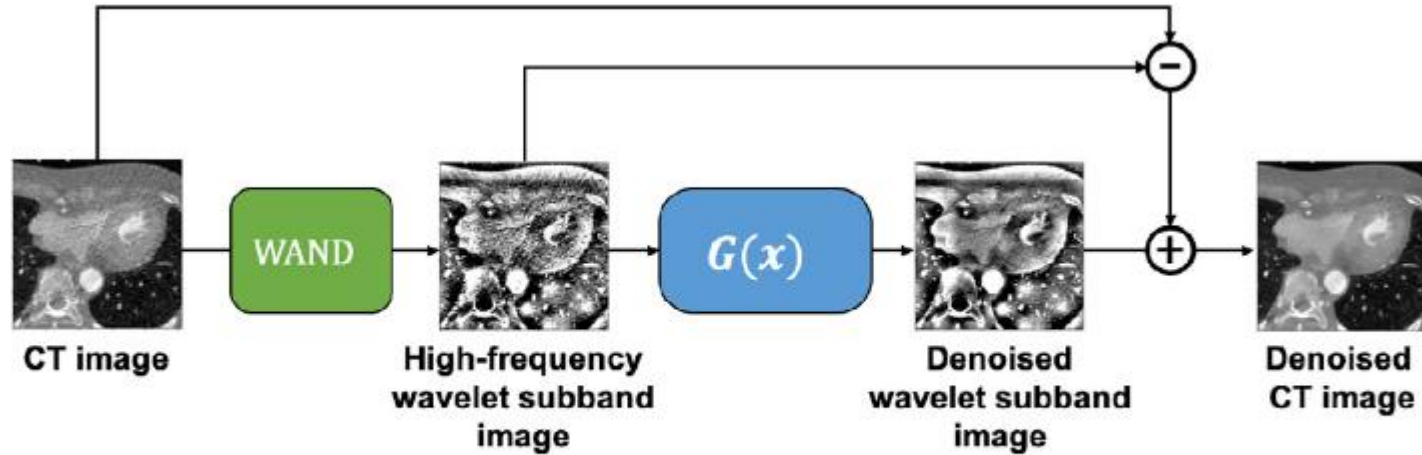


Fig. 6. Inference step of the proposed denoising network. WAND: wavelet-assisted noise disentanglement block.

Results

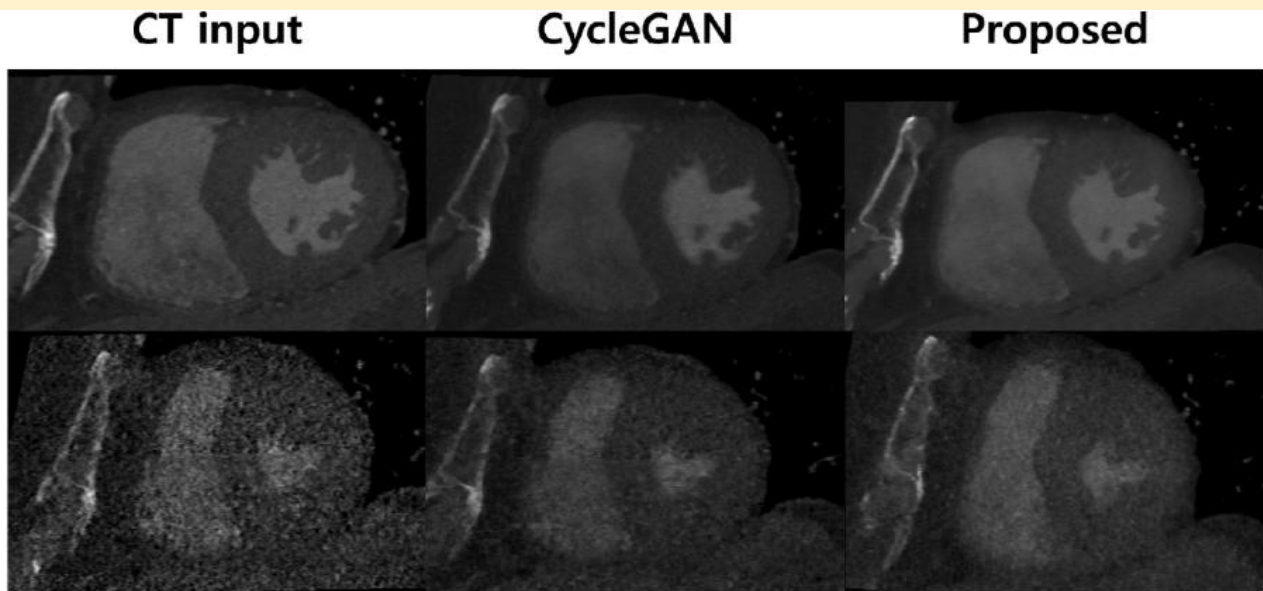


Fig. 15. CT image. From the left to right column, the images are from the CT input, the cycleGAN method (Kang et al., 2019), and the proposed method. The upper row are the normal-dose CT images in the diastolic phase. The lower row are the low-dose CT images in the systolic phase. The intensity range of the CT image is (1275, 1725) [HU].

Denoising diffusion wavelet models for zero-shot medical image translation

- Task: Zero-Shot CBCT to CT translation
- Proposed Model: Denoising Diffusion Wavelet Models (DDWM)
- Core Concept - “Similarity Bridge”
 - As shown in Fig.4
 - Y represents the shared information(anatomy / structure)
 - X and Z represent the Domain Differences(Noise / Texture)
 - In medical image: the low-frequency wavelet coefficients act as Y(the bridge)

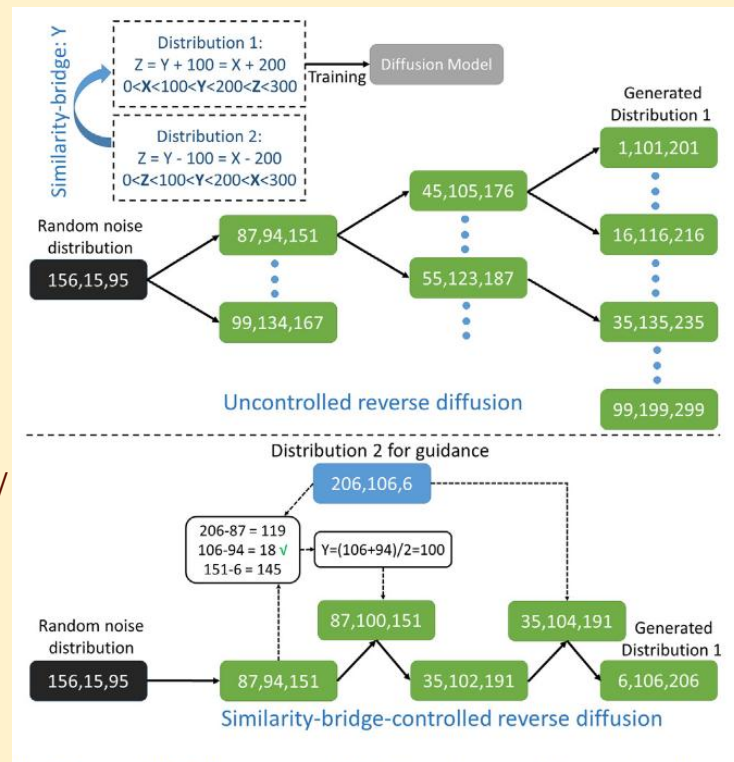
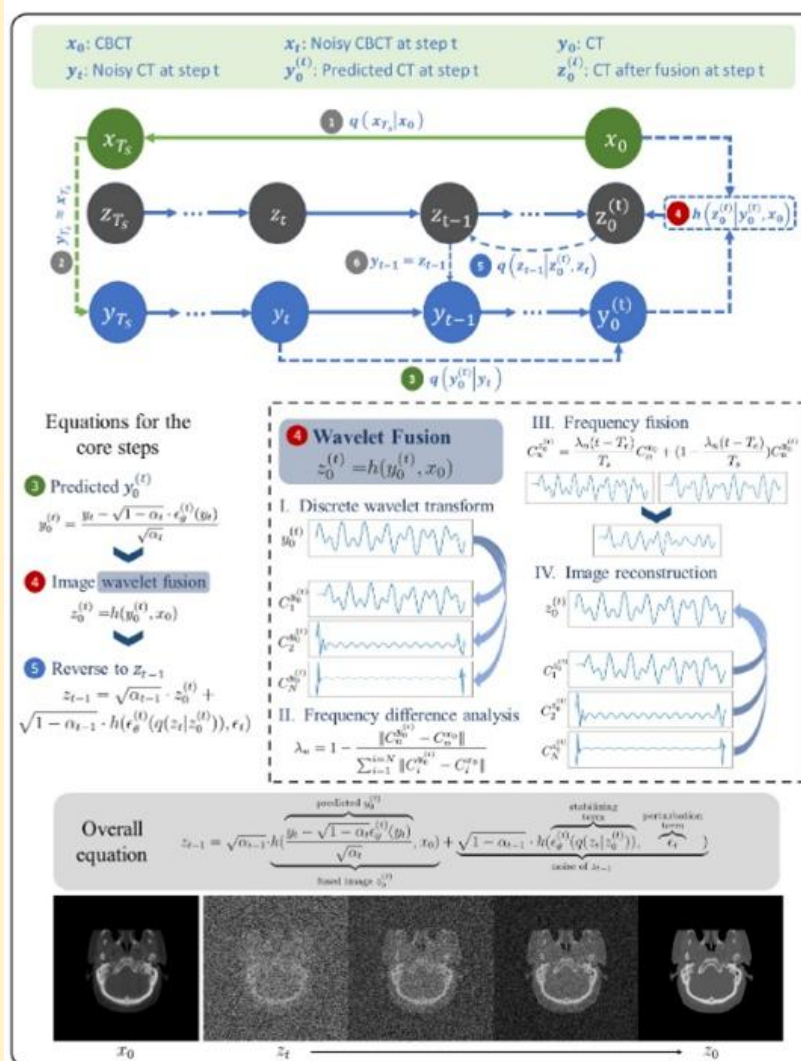


Fig. 4. A simplified illustration explaining the working principle of the similarity-bridge-controlled reverse diffusion process using an array-based generative model. In this example, Y acts as a similarity bridge between two distributions, where Distribution 2 provides guidance during the generation of Distribution 1 by the diffusion model.

The framework of DDWM

- Instead of starting from pure Gaussian noise, DDWM start from the Forward-diffused CBCT
- Adaptive Wavelet Fusion(step 4)
 - Low-frequency: retain information from source CBCT
 - High-frequency: rely on diffusion model prediction
 - Apply IDWT to reconstruct the fused image for the next t-1



Experimental Results: Zero-shot & OOD robustness

- DDWM was evaluated on OOD dataset, the results show that DDWM maintains high performance with the different data distribution
- By locking the low-frequency(structure) components via Wavelet Fusion, the model is forced to respect the anatomical structure, preventing from generating fake details

Table 2

Comparison of different methods on out-of-distribution Dataset II. Downward arrows indicate “lower is better” and upward arrows indicate “higher is better”. The values in parentheses are the p-values obtained from performing the Wilcoxon signed-rank test with our DDWM.

Method	FID ↓	PSNR ↑	DICE ↑	MAE (HU) ↓
CycleGAN	98.8879	28.5057 ± 2.2388 (1.05e−128)	0.7855 ± 0.2239 (7.00e−17)	36.9698 ± 13.8119 (2.02e−157)
GcGAN	87.1466	29.0956 ± 2.5937 (1.30e−29)	0.7817 ± 0.2161 (7.57e−31)	36.8642 ± 14.0014 (1.90e−141)
RegGAN	135.4764	23.0311 ± 3.8592 (1.31e−169)	0.6302 ± 0.2596 (4.86e−156)	90.2041 ± 76.0190 (1.31e−169)
UNIT	86.7005	26.1389 ± 2.0831 (3.42e−163)	0.6776 ± 0.1900 (4.79e−155)	50.3690 ± 17.3299 (2.65e−166)
MUNIT	104.4530	26.2132 ± 2.2055 (1.15e−165)	0.7196 ± 0.2088 (2.82e−145)	49.4790 ± 17.4753 (3.28e−167)
SDEdit	115.9626	23.2992 ± 2.2013 (1.33e−169)	0.5429 ± 0.2447 (1.03e−156)	65.6011 ± 19.9472 (2.38e−169)
SynDiff	91.0701	24.4304 ± 2.1566 (1.94e−168)	0.7147 ± 0.2185 (1.93e−138)	56.7172 ± 19.6626 (1.80e−169)
EGSDE	127.7572	27.2243 ± 2.3557 (2.85e−168)	0.6822 ± 0.2212 (1.22e−151)	45.3046 ± 15.4370 (2.25e−168)
FGDM	85.6288	28.8324 ± 2.2865 (2.25e−149)	0.7748 ± 0.2126 (5.08e−91)	35.4334 ± 10.9221 (2.33e−149)
MIDiffusion	131.8422	26.7871 ± 2.1890 (3.18e−168)	0.7092 ± 0.2178 (7.25e−146)	44.1551 ± 14.3066 (2.19e−168)
DDWM	61.0412	29.9331 ± 2.1610	0.8065 ± 0.2136	28.9325 ± 9.4438

Ablation study: The role of frequency components

- **Case A: without low-frequency fusion**
 - The overall shape and anatomical geometry are distorted
 - The model loses its structural anchor
- **Case B: without high-frequency fusion**
 - The model lacks the necessary texture guidance
- **Conclusion: both components are essential for the complete DDWM performance**

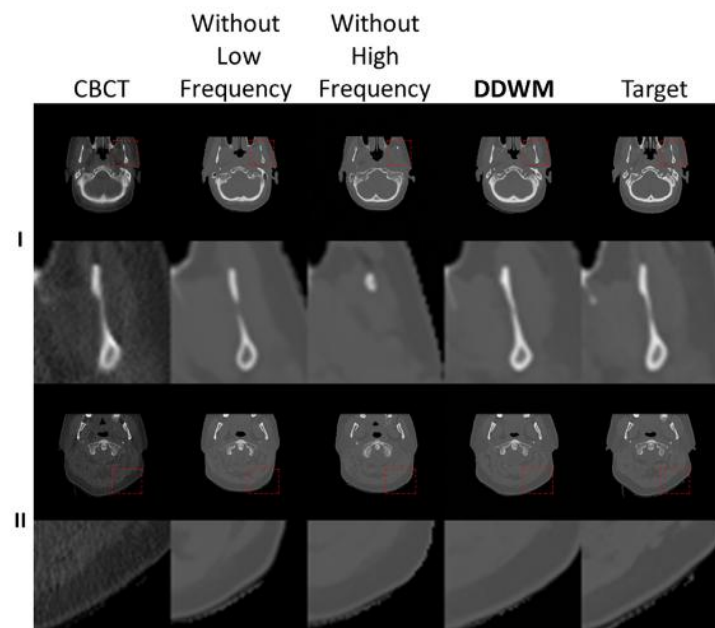


Fig. 11. Visual comparison showing the effect of low-frequency and high-frequency components.

Ablation study: Wavelet stopping steps T_e

- T_e is the time step where we stop performing Wavelet Fusion and let the standard Diffusion process take over

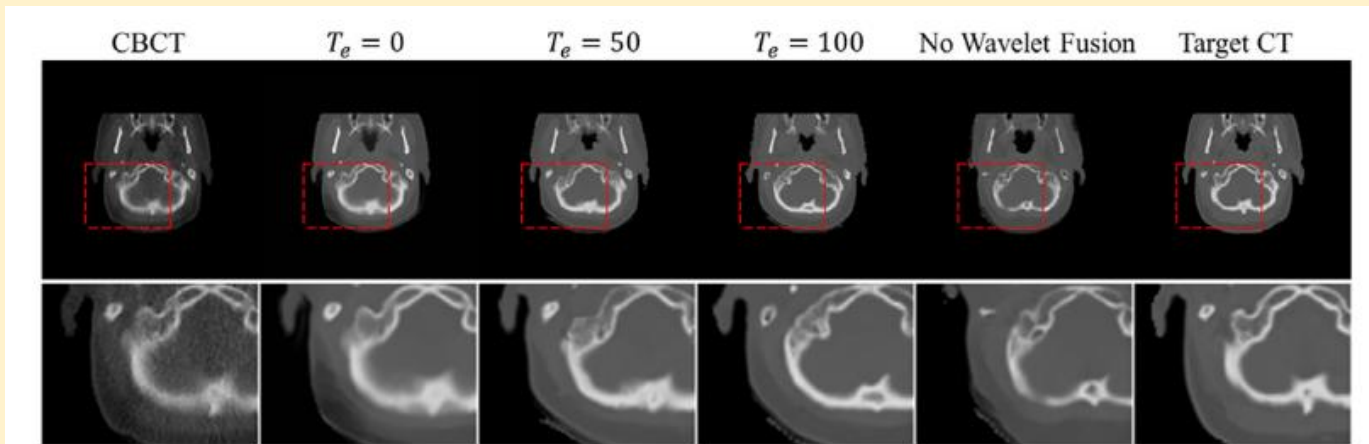


Fig. 9. Visualization of the generated CT y_0 under the condition of $T_e = 0, 50, 100$, and without wavelet fusion.

Conclusion

- Limitations:
 - Inference efficiency
 - The similarity bridge relies on the assumption that source and target share similar underlying information, future work might expand the framework to distinct modalities, where the frequency domain relationships are more complicated
- Traditional models try to learn everything at once, therefore, DDWM strategy decouples the problem into target domain and generation control, while the later is achieved by Adaptive Wavelet Fusion
- By integrating frequency constraints, we bridge the gap between data-driven creativity and physical reality, making generative models safe for clinical adoption.

References

- Gu, J., Yang, T. S., Ye, J. C., & Yang, D. H. (2021). CycleGAN denoising of extreme low-dose cardiac CT using wavelet-assisted noise disentanglement. *Medical image analysis*, 74, 102209.
- Song, J., Jeong, J. H., Park, D. S., Kim, H. H., Seo, D. C., & Ye, J. C. (2020). Unsupervised denoising for satellite imagery using wavelet directional CycleGAN. *IEEE Transactions on Geoscience and Remote Sensing*, 59(8), 6823-6839.
- Li, Y., Kong, X., Xie, J., Ver Steeg, G., & Zhang, Y. (2025). Denoising diffusion wavelet models for zero-shot medical image translation. *Knowledge-Based Systems*, 113800.