Department of Medical Physics
UNIVERSITY OF WISCONSIN
SCHOOL OF MEDICINE AND PUBLIC HEALTH
DARTMOUTH ENGINEERING

# Denoising for High Fidelity External Beam Therapy Video Cherenkov Images Using a Deep Learning Methodology

Jeremy Hallett[1,4], Shiru Wang[2], Yucheng Tang[5], Yujia Wu[2], Alexander Geiersbach[2], Daguang Xu, Lesley A. Jarvis[3], Petr Bruza[2,4], Brian W. Pogue[1,2,4]

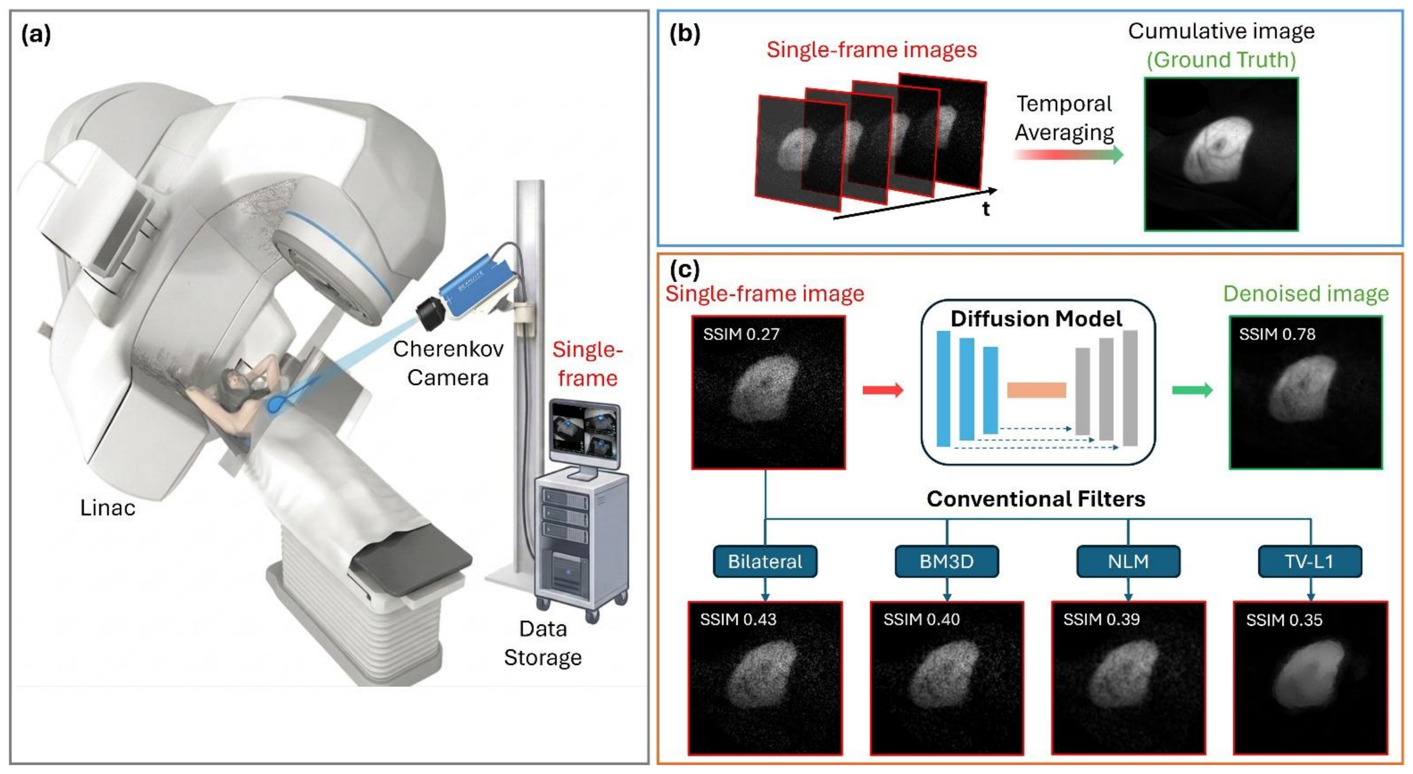
[1] University of Wisconsin-Madison, Madison, WI [2] Dartmouth College, Hanover NH [3]Dartmouth Cancer Center [4]DoseOptics LLC, Lebanon NH [4]NVIDIA Corp, Santa Clara CA,

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING
MEDICAL PHYSICS COMMUNITY INSPIRATION INNOVATION IMPACT

## Motivation & Background

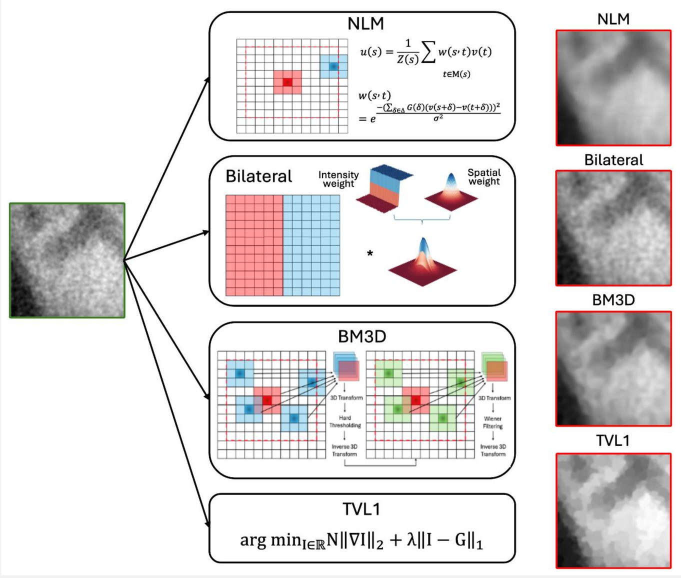
### 1 Cherenkov Imaging

Cherenkov light is produced in tissue from high energy electrons that are released during the dose delivery process of radiation therapy. This light can be used to visualize the radiotherapy beam and detect medical incidents. The Cherenkov images are captured through time-sequenced intensifier gating of the camera to the linac pulses [1]. High fidelity images are generated through the summation of many image frames. Single frame images, however, exhibit extreme noise levels




Overview of Cherenkov imaging and denoising pipeline. (a) Schematic of the intensified CMOS camera system capturing Cherenkov emissions during radiation therapy. (b) Temporal averaging of consecutive single-frame acquisitions to produce a high-SNR cumulative image. (c) Comparison of conventional denoising methods (NLM, BM3D, bilateral filter, TV-L1) and proposed deep learning approaches applied to a single Cherenkov frame.

### 2 Classical Denoising Algorithms




While algorithmic approaches have been explored for cumulative image denoising [3], the noise exhibited in single frame images, limits dose interpretability for video sequences. This is due to several factors:

- External beam generated Cherenkov light as a fluence rate of 0.3μW/cm² [1,4]
- short acquisition time allowing for video rate imaging

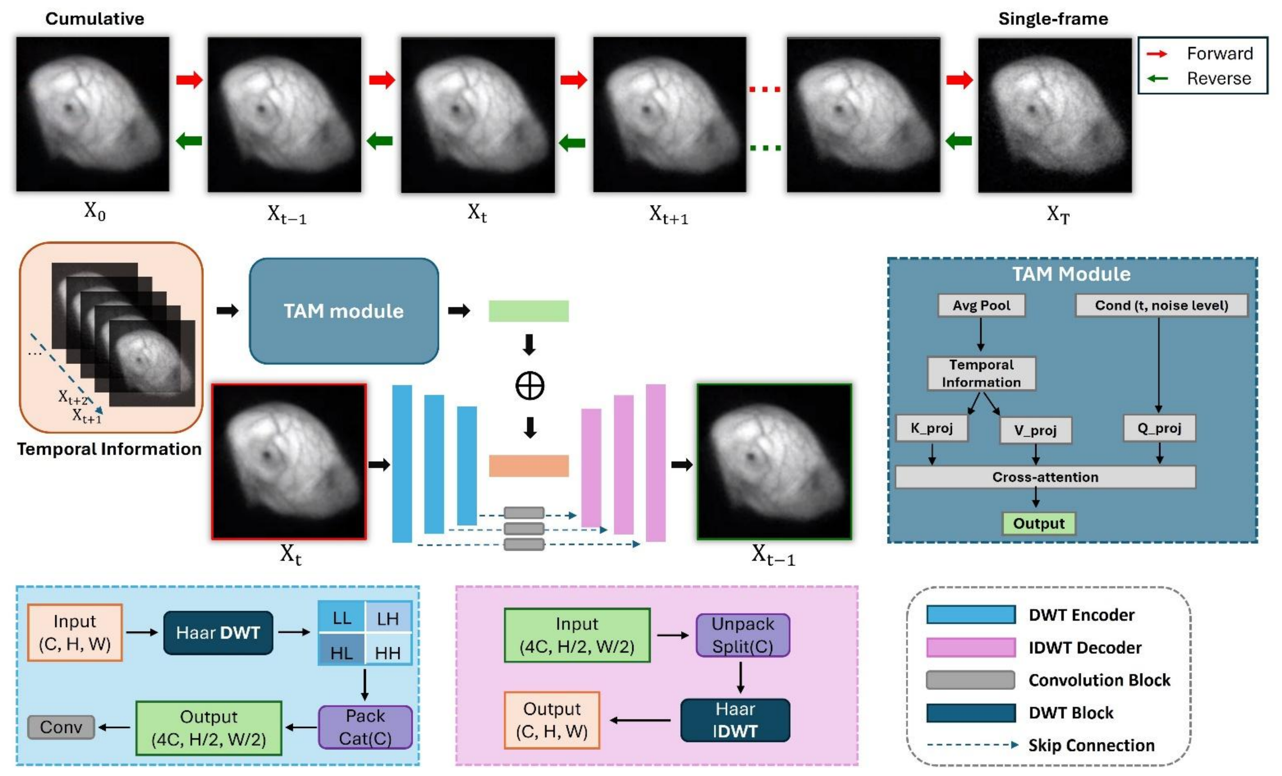
Deep Learning has become the gold standard for medical imaging applications and is explored further in this study

## Methods & Models

### 1 Wavelet Diffusion Model

Several machine learning denoisers were tested such as GAN and a classic diffusion denoiser. Additionally, a more complex approach was proposed as shown to the right with the following features:

- Conditional Diffusion Network: Every noise removal step of a gaussian noise image is conditioned with the original noisy Cherenkov image [5]
- Temporal Attention Module (TAM): Logs bottleneck features from previously denoised images to inform future frames for temporal consistency
- Denoising Diffusion Implicit Model sampling (DDIM): Allows for larger denoising steps to improve efficiency [6]
- Wavelet Encoder-Decoder: Frequency components are decomposed using the Haar wavelet before denoising weights are applied




Architecture of the proposed conditional temporal-diffusion network. The model is built using a standard diffusion method with a wavelet decoder network taking the place of the standard UNET. The TAM module is designed to maintain temporal consistency.

### 2 Training/Testing Datasets

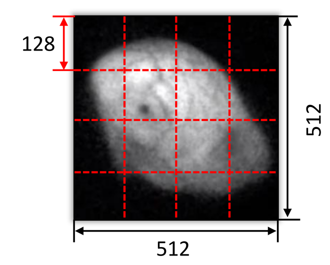
**Algorithm 1** Cherenkov Noise Model

**Input:** Cumulative clean image $x \in [0,1]^{H\times W}$; illumination level $I_0$; integration time $t_{int}$; pixel pitch $\delta$; gain $\boldsymbol{G}$; blur parameters $(\sigma, k)$; read out noise $\sigma_r$; intensity noise scale $\alpha$; number of frames $\boldsymbol{N}$

**Output:** Noisy frame stack $\{y_1,\ldots,y_N\}$ ; cumulative images $\{\bar{y_c}\}$ for selected frame counts, where $c \in C$

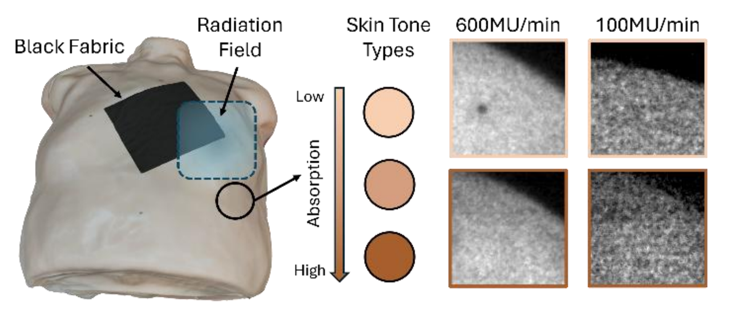
1. Compute mean photon map: $\mu \leftarrow x \cdot I_0 \cdot t_{int} \cdot \delta^2$
2. **for** i = 1 to $N$ **do**
3. Sample photoelectrons: $e \leftarrow Poisson(\mu)$
4. Apply gain: $s \leftarrow \boldsymbol{G} \cdot e$
5. Apply intensifier blur: $s_b \leftarrow GaussianBlur(s,\sigma,k)$
6. Normalize blurred signal: $\bar{s_b} \leftarrow s_b/max(s_b)$
7. Compute noise scale map: $\beta \leftarrow 1+\alpha\cdot\bar{s_b}$
8. Sample read noise: $n \leftarrow \mathcal{N}(0,\sigma_r)\odot\beta$
9. Digitize: $y_i \leftarrow \lfloor s_b + n \rfloor$
10. **end for**
11. **for** each $c \in C$ **do**
12. Compute cumulative image: $\bar{y_c} \leftarrow \frac{1}{c}\sum_{i=1}^{c} y_i$
13. **end for**
14. **return** $\{y_1,\ldots,y_N\}$, $\{\bar{y_c}\}$

- Training Set: Simulated single frame images using the algorithm outlined to the left. Cumulative clinical Cherenkov images were selected and run through the simulation to generate the entire training set
- The simulation estimates a photon map using tunable pixel pitch, illumination, integration time parameters. Single frame images can then be generated by sampling this map and tuning the gain, blur and read noise of the sample




- Images were split into 128x128 patches, A400 GPU, 1000 epochs, AdamW, cosine annealed learning rate (1e-4->1e-5)

- Clinical Dataset: Raw, single frame images from outdated camera with excessive noise
- Phantom images generated with current generation camera. Black fabric used to create sharp edges for image sharpness metrics




Anthropomorphic tissue-mimic phantom with interchangeable skin tone layers used for model evaluation. Single-frame Cherenkov images of the lightest and darkest skin tones are shown on the right, illustrating the variation in signal intensity and noise characteristics across different pigmentation conditions.

## Results

1

| Method | SSIM↑ | PSNR↑ | \|Residual mean\|↓ | Temporal SSIM↑ |
|---|---|---|---|---|
| Noise input | 0.74 ± 0.09 | 27 ± 4 | 0.018 ± 0.010 | 0.86 ± 0.02 |
| NLM | 0.81 ± 0.15 | 19 ± 5 | 0.10 ± 0.06 | 0.98 ± 0.02 |
| U-Net | 0.82 ± 0.09 | 28 ± 4 | 0.018 ± 0.010 | 0.92 ± 0.01 |
| Condition GAN | 0.80 ± 0.05 | **29 ± 5** | 0.017 ± 0.010 | 0.906 ± 0.007 |
| Vanilla Diffusion | 0.79 ± 0.05 | 28 ± 4 | **0.013 ± 0.010** | 0.91 ± 0.01 |
| TAM-Freq Diffusion | 0.82 ± 0.09 | 28 ± 4 | 0.019 ± 0.010 | **0.92 ± 0.01** |
| Wavelet Diffusion | **0.83 ± 0.07** | 28 ± 4 | 0.016 ± 0.010 | **0.92 ± 0.01** |

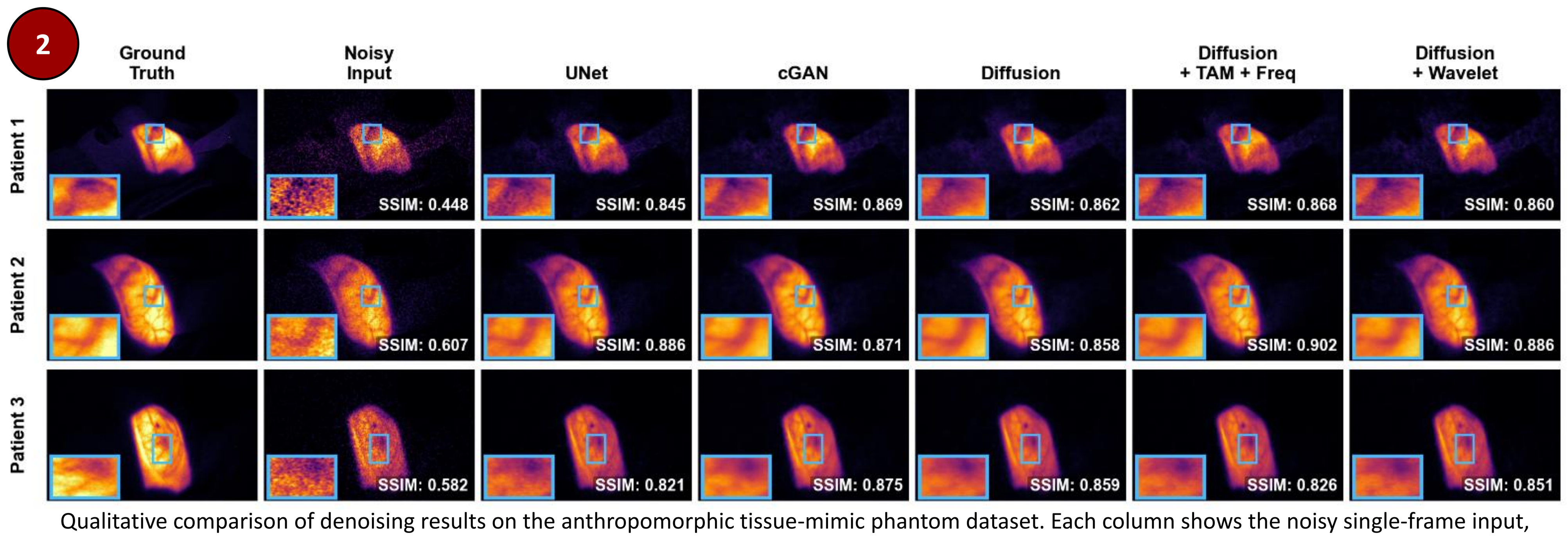
Quantitative model performances on phantom data.

| Method | SSIM↑ | PSNR↑ | \|Residual mean\|↓ | Temporal SSIM↑ |
|---|---|---|---|---|
| Noise input | 0.47 ± 0.12 | 25 ± 4 | 0.018 ± 0.013 | 0.48 ± 0.12 |
| NLM | 0.77 ± 0.11 | 20 ± 6 | 0.103 ± 0.064 | 0.84 ± 0.09 |
| Unet | 0.81 ± 0.07 | 29 ± 6 | 0.018 ± 0.013 | 0.84 ± 0.04 |
| Condition GAN | **0.82 ± 0.05** | **30 ± 6** | 0.014 ± 0.012 | 0.88 ± 0.02 |
| Vanilla Diffusion | **0.82 ± 0.04** | 29 ± 6 | **0.013 ± 0.011** | 0.88 ± 0.01 |
| TAM-Freq Diffusion | **0.82 ± 0.07** | 29 ± 6 | 0.019 ± 0.013 | 0.88 ± 0.02 |
| Wavelet Diffusion | **0.82 ± 0.06** | 29 ± 6 | 0.017 ± 0.013 | **0.89 ± 0.03** |

Quantitative model performances on clinical patient data.

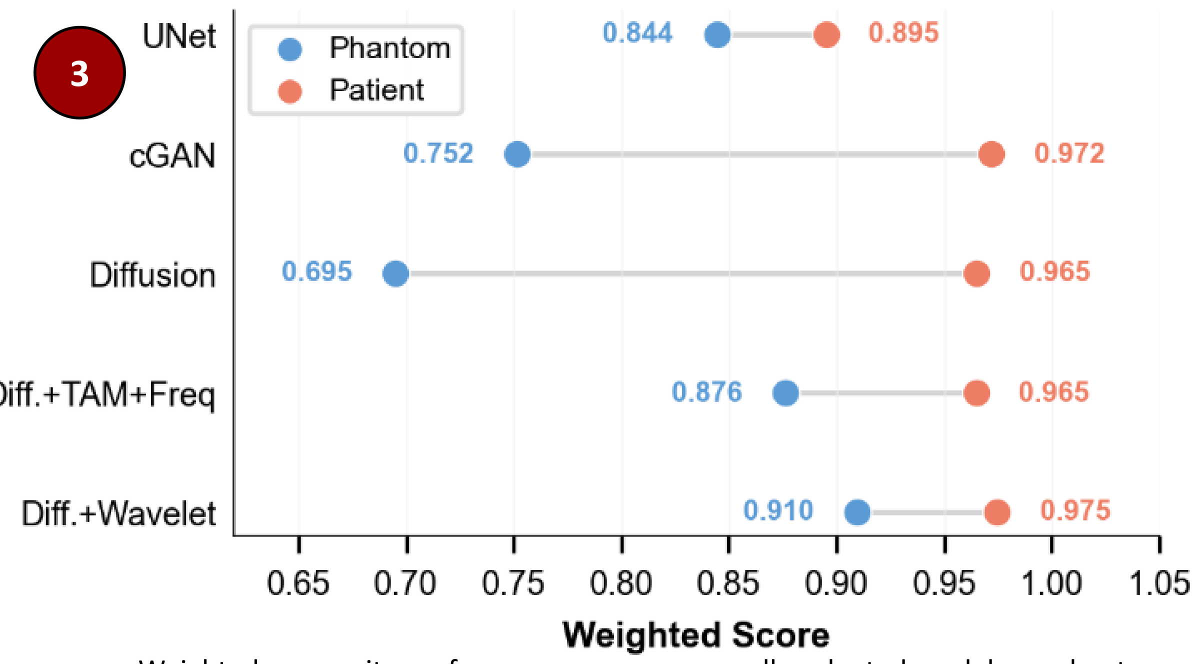
1) These tables demonstrate that some models are more ideal for specific outcomes than others (i.e. the proposed wavelet diffusion performs best for similarity to the gold standard)
2) The resulting contrast of each denoised image is more similar to the contrast exhibited in the initial single frame image than what is present in the cumulative image
3) Across both the phantom and patient data, the diffusion wavelet model performs best when considering a weighting of all metrics.
4) 3 DDIM steps was determined to produce sufficient denoising while allowing for efficient computation
5) Image sharpness is sacrificed when performing machine learning denoising. Less complex models (UNET) lead to higher less blurring but undeforms the more complex models on denoising

2



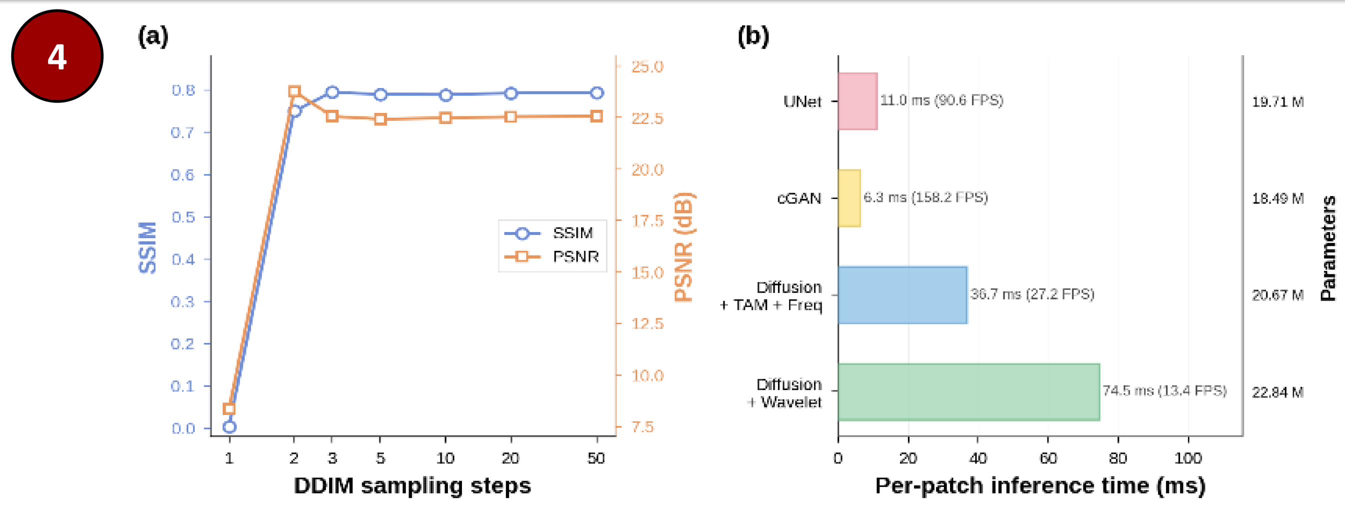

Qualitative comparison of denoising results on the anthropomorphic tissue-mimic phantom dataset. Each column shows the noisy single-frame input, cumulative ground truth, and outputs from all evaluated methods, with zoom-in ROIs highlighting structural detail preservation.

3



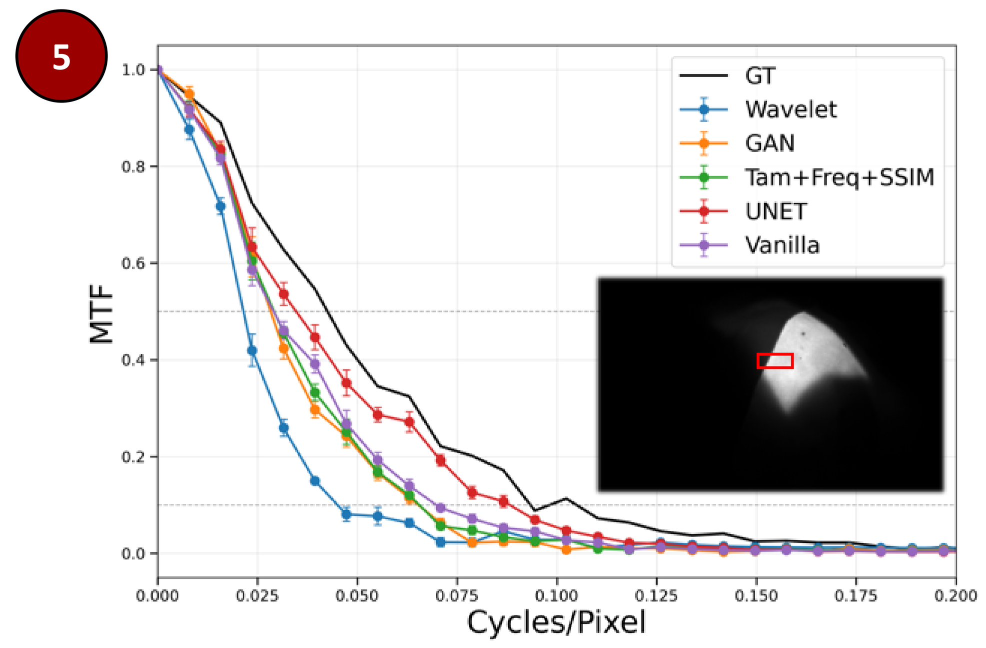

Weighted composite performance scores across all evaluated models on phantom and patient datasets. We compute the weighted composite score with weights assigned as 0.35, 0.50 and 0.15 for SSIM, Temporal SSIM, and PSNR

4




Computational efficiency analysis. (a) Effect of the number of DDIM sampling steps on output quality, measured by SSIM and PSNR, demonstrating rapid convergence at 2–3 steps. (b) Per-patch inference time and total parameter count for all evaluated models, measured on an NVIDIA RTX A4000 GPU with 128 × 128 input patches.

5




| Model | MTF50% | MTF10% |
|---|---|---|
| Wavelet Diffusion | 0.022 | 0.045 |
| GAN | 0.028 | 0.065 |
| TAM+Freq+SSIM | 0.029 | 0.066 |
| UNET | 0.035 | 0.088 |
| Vanilla Diffusion | 0.029 | 0.070 |

MTF plot of a phantom imaged denoised using the various denoising methods. The table lists the 50% and 10% crossings of the MTF plots with larger values corresponding to better sharp edge preservation.

## References

[1] Parks, A., Hallett, J., et al., *Review of Cherenkov Imaging technology advances in radiotherapy: single-photon-level imaging in high ambient light and radiation backgrounds.* Biophotonics Discovery, 2024.
[2] Hallett, J.E., M. Jermyn, and B.W. Pogue, *Deep learning model for Cherenkov image de-identification in radiotherapy delivery verification.* Medical Imaging 2026: Image Perception, Observer Performance, and Technology Assessment (Proceedings), 2026. **13928**.
[3] Hallett, J. E., et al. (2024). "Noise & mottle suppression methods for cumulative Cherenkov images of radiation therapy delivery." Phys Med Biol **69(22).**
[4] Glaser, A. K., et al. (2015). "Cherenkov radiation fluence estimates in tissue for molecular imaging and therapy applications." Phys Med Biol **60(17): 6701-6718.**
[5] Saharia, C., et al., *Palette: Image-to-Image Diffusion Models.* arXiv preprint, 2022.
[6] Song, J., et al. (2021). "DENOISING DIFFUSION IMPLICIT MODELS." ICLR.

## Acknowledgements/Disclosure

This work was partially supported by DoseOptics LLC

## Contact Information

jehallett@wisc.edu