



DWTGS: Rethinking Frequency Regularization for Sparse-view 3D Gaussian Splatting

Hung Nguyen, Runfa Li, An Le, Truong Nguyen

Video Processing Lab, UC San Diego
 {hun004, ru1002, d01e, tqn001}@ucsd.edu



Motivation

- 3D Gaussian Splatting (3DGS) represents 3D scenes with optimizable 3D Gaussians, reconstructed from a collection of multi-view 2D images.
- However, 3DGS traditionally relies on dense training views with accurate camera poses. Under sparse-view settings, the reconstructed 3D geometry is inadequately constrained, thus severely compromising rendering quality.
- We introduce **DWTGS**, a framework that leverages the **Discrete Wavelet Transform** at a loss function level to regularize sparse-view 3DGS.

Discrete Wavelet Transform (DWT)

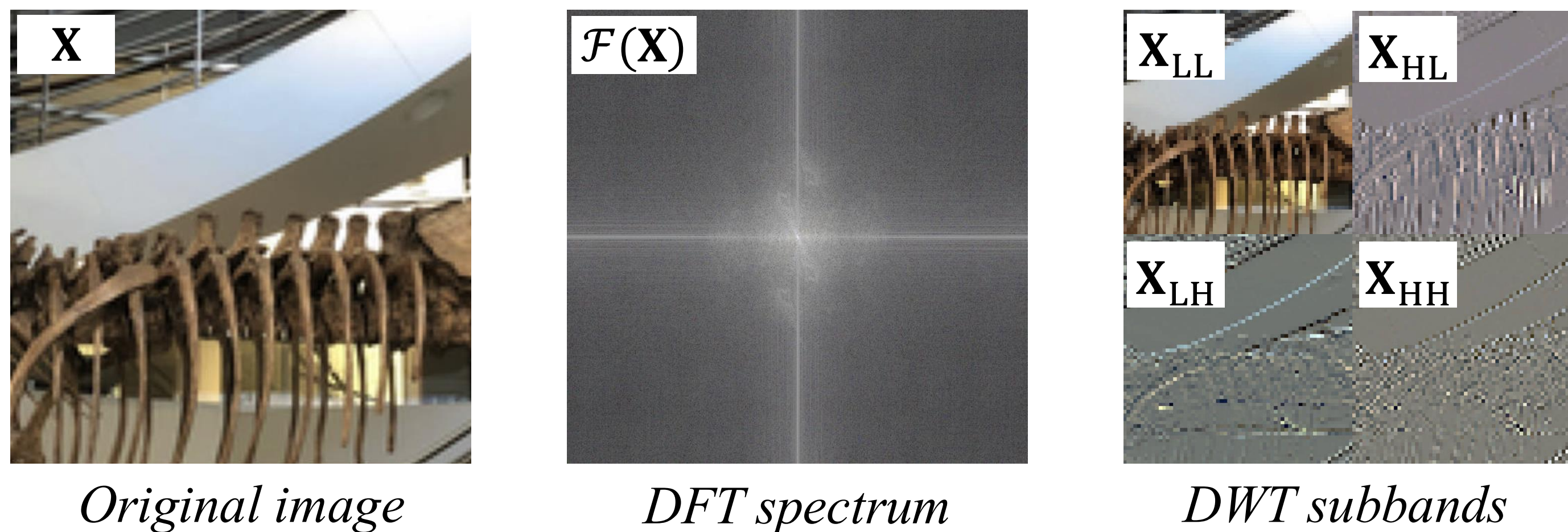
- The DWT decomposes an input image $\mathbf{X}$ into varying levels of frequencies, called “subbands”. It can be implemented using matrix multiplication with low-pass and high-pass filters $\mathbf{L}$ and $\mathbf{H}$:

$$\mathbf{X}_{LL} = \mathbf{L}\mathbf{X}\mathbf{L}^T, \mathbf{X}_{LH} = \mathbf{H}\mathbf{X}\mathbf{L}^T, \mathbf{X}_{HL} = \mathbf{L}\mathbf{X}\mathbf{H}^T, \mathbf{X}_{HH} = \mathbf{H}\mathbf{X}\mathbf{H}^T$$

where $\mathbf{L}$ and $\mathbf{H}$ are Toeplitz matrices constructed from 1D analysis filters ℓ and h , respectively. In the Haar case:

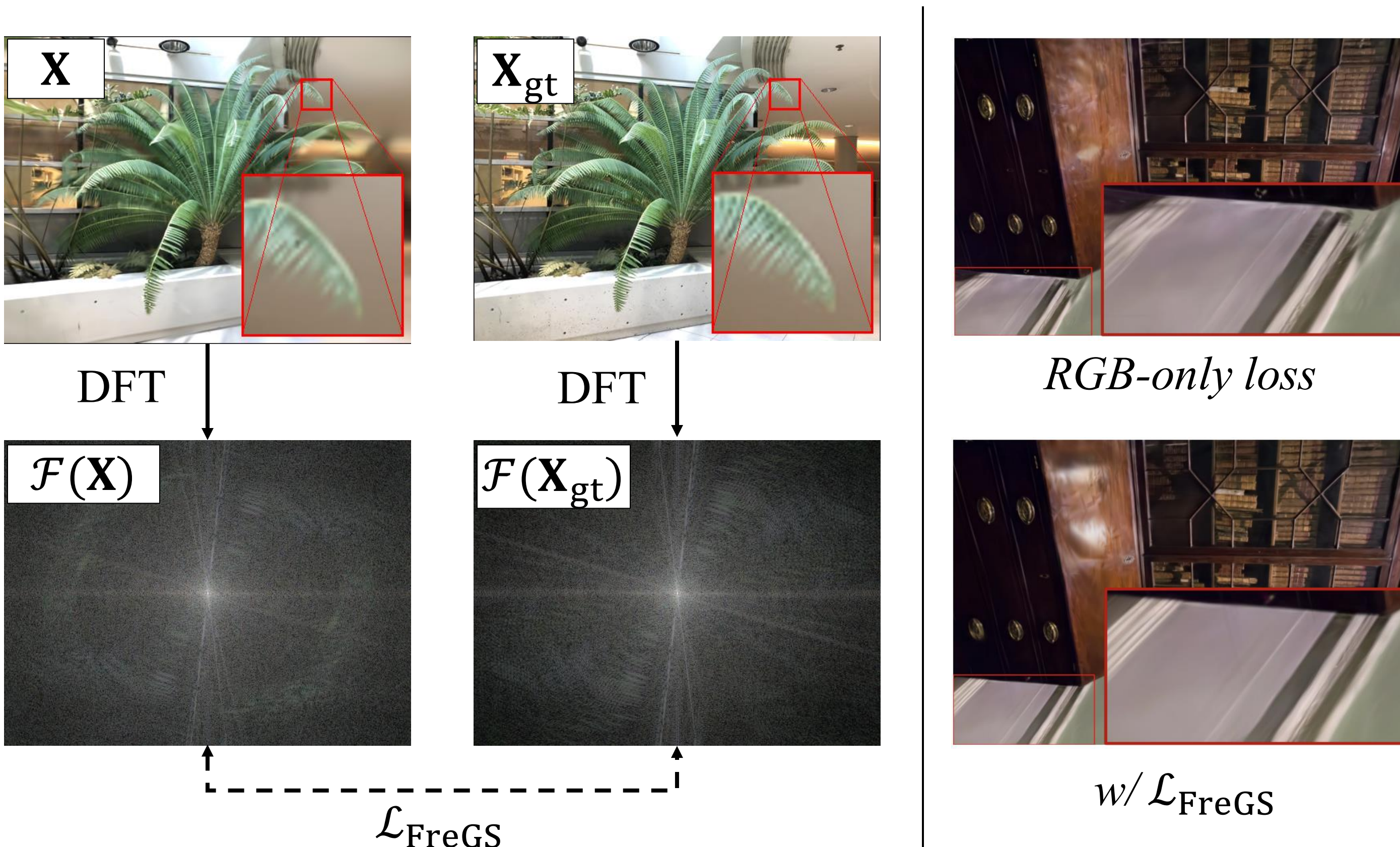
$$\ell = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right], h = \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right]$$

- Compared to the Discrete Fourier Transform (DFT), the DWT represents a spatially interpretable form of frequency decomposition.



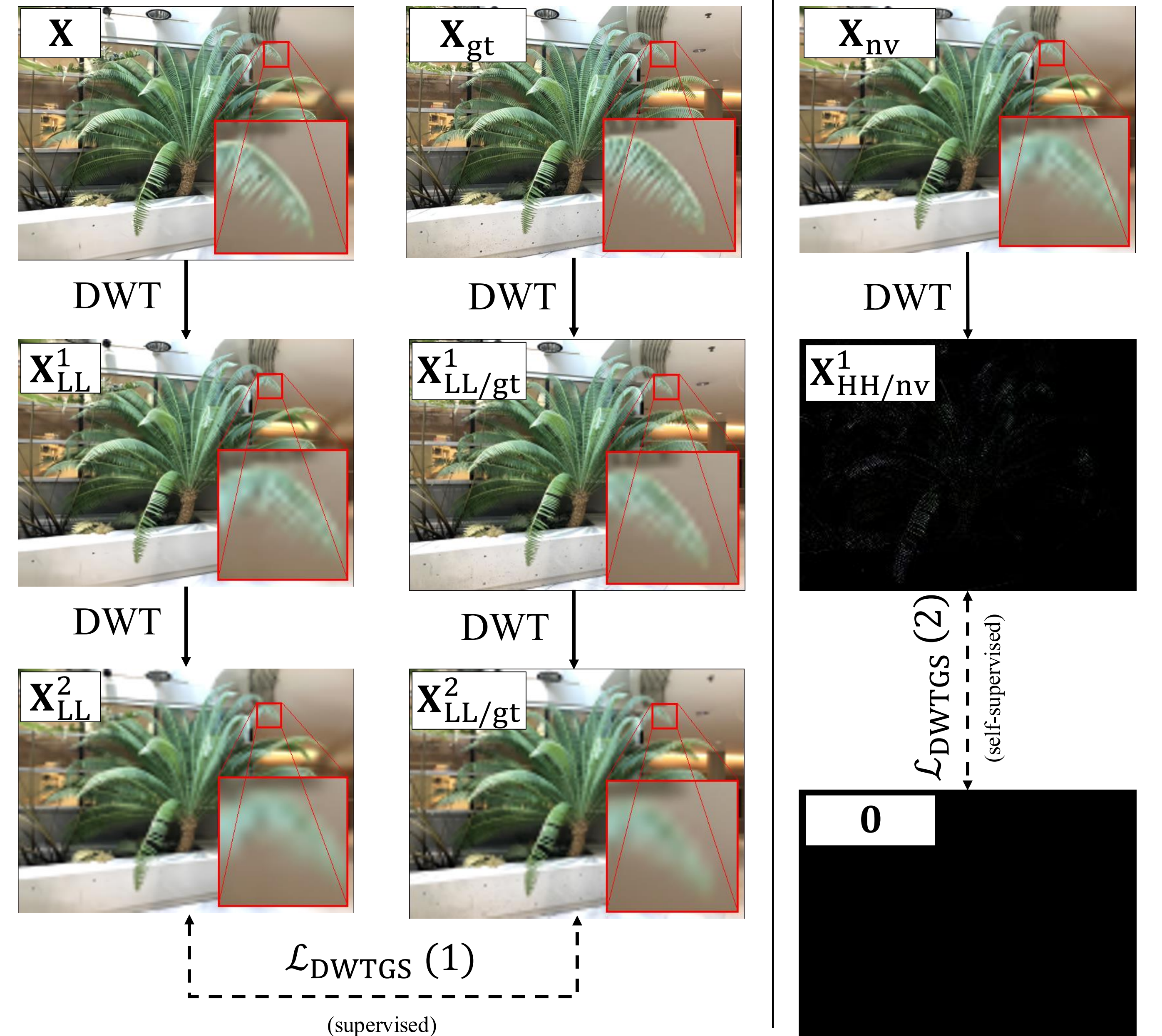
Frequency Regularization for 3DGS

- For dense-view settings, FreGS [1] leverages an auxiliary loss, $\mathcal{L}_{\text{FreGS}}$, between DFT spectra of ground-truth and rendered images, thus mitigating the low-frequency bias of RGB-space supervision.



Methodology

- Based on the DFT, $\mathcal{L}_{\text{FreGS}}$ needs manual balancing of low-/high-frequency (LF/HF) components and magnitude/phase terms, as well as specifying additional filtering operations.
- We introduce $\mathcal{L}_{\text{DWTGS}}$, a wavelet-domain loss that offers readily available and interpretable LF/HF supervision.

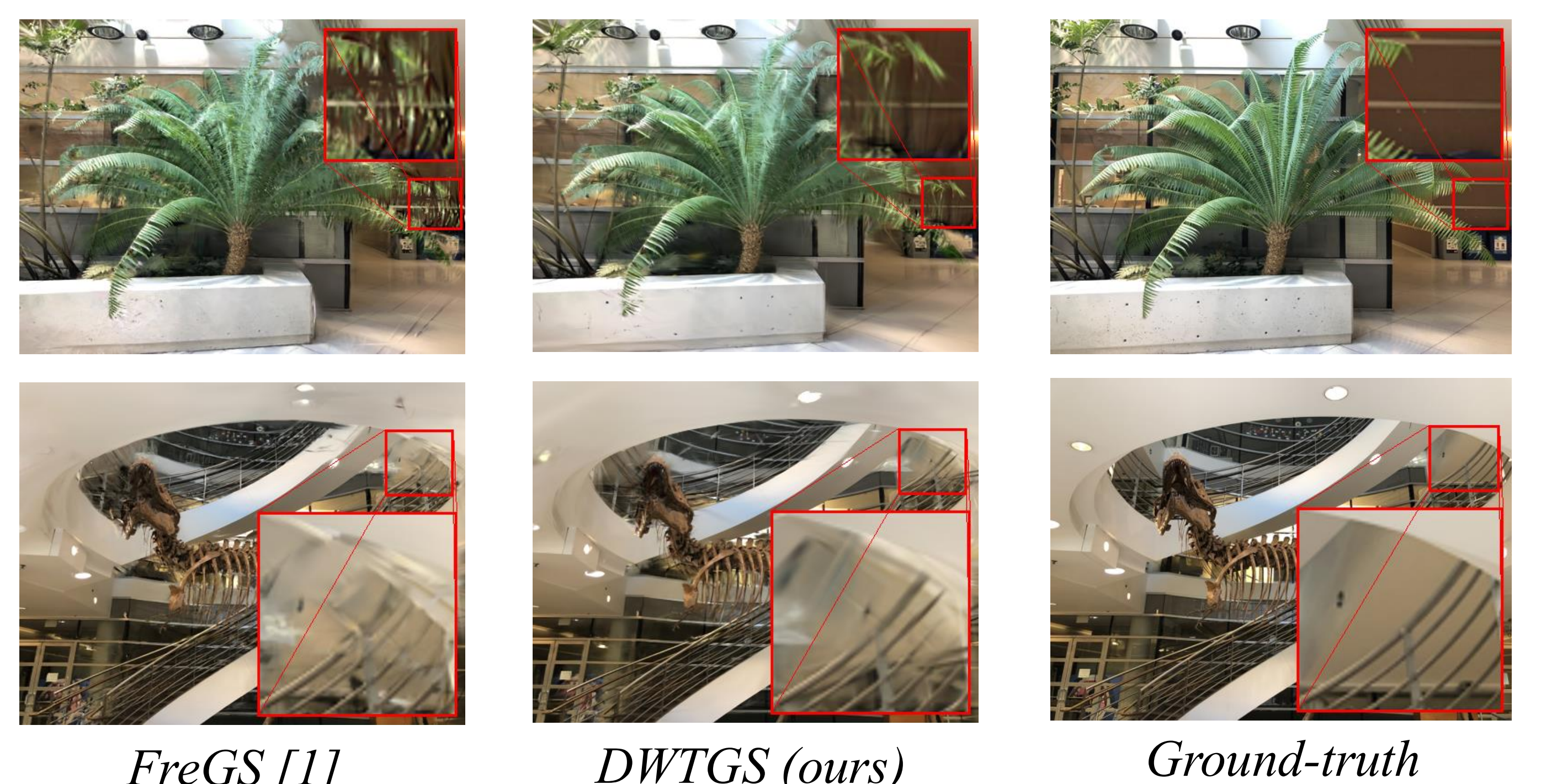


- $\mathcal{L}_{\text{DWTGS}}$ includes:

- a supervised term (1) between the ground-truth and rendered LL subbands
- a self-supervised term (2) that enforces the HH subband of novel renders (without ground-truths) to be zero.
- This effectively mitigates HF hallucinations, which in sparse-view settings are caused by unreliable supervision of HF structures.

Qualitative & Quantitative Results

	3-view LLFF dataset [2]			12-view Mip-NeRF 360 dataset [3]		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
3DGS	20.39	0.706	0.198	19.30	0.564	0.352
FreGS	20.40	0.706	0.198	19.41	0.571	0.344
DWTGS (ours)	20.78	0.720	0.197	19.75	0.590	0.343



References

- [1] Zhang et al. (2024). FreGS: 3D Gaussian Splatting with Progressive Frequency Regularization. CVPR.
- [2] Mildenhall et al. (2019). Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines. TOG.
- [3] Barron et al. (2022). Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. CVPR.