



From Coarse to Fine: Learnable Discrete Wavelet Transforms for Efficient 3D Gaussian Splatting

Hung Nguyen, An Le, Runfa Li, Truong Nguyen
Video Processing Lab, UC San Diego



Motivation

- 3D Gaussian Splatting (3DGS) represents 3D scenes with optimizable 3D Gaussians, reconstructed from a collection of multi-view 2D images.
- However, Gaussian counts often grow uncontrollably during training. Therefore, reducing them, while preserving rendering quality, is crucial to cut GPU memory, storage costs, and improve efficiency.
- We introduce **AutoOpti3DGS**, a framework for **automatically optimizing 3DGS** Gaussian counts by modulating the frequencies of the input images with the Learnable Discrete Wavelet Transforms (DWTs).

Discrete Wavelet Transforms

- The Forward DWT decomposes an input image into varying levels of frequencies, called “subbands”. It can be implemented using matrix multiplication with low-pass and high-pass *analysis* filters **L** and **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 **L** and **H**, using the Haar wavelet, can be of the following form:

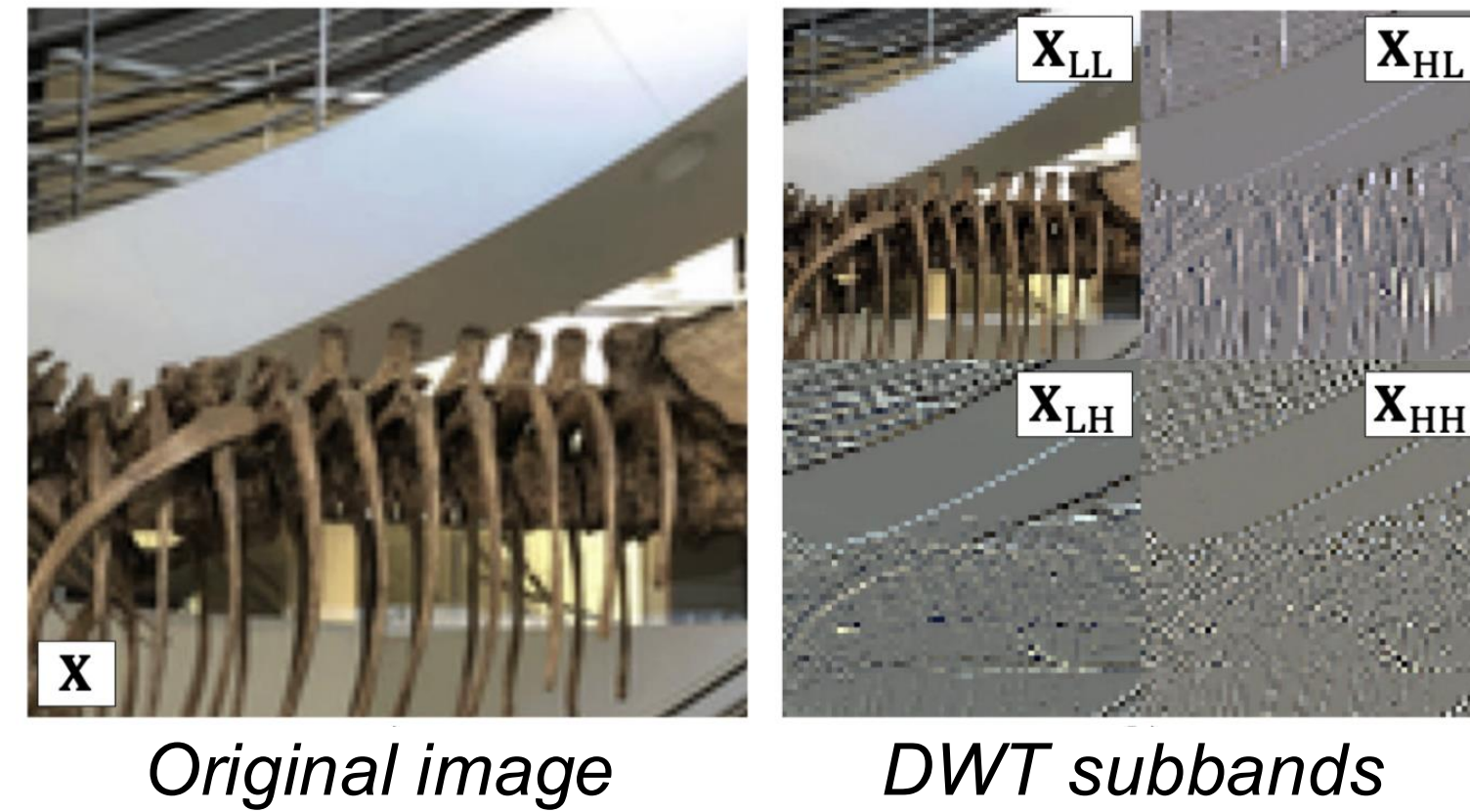
$$\mathbf{L} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & \dots \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}, \mathbf{H} = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & \dots \\ 0 & 0 & -1/\sqrt{2} & 1/\sqrt{2} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

(low-pass) (high-pass)

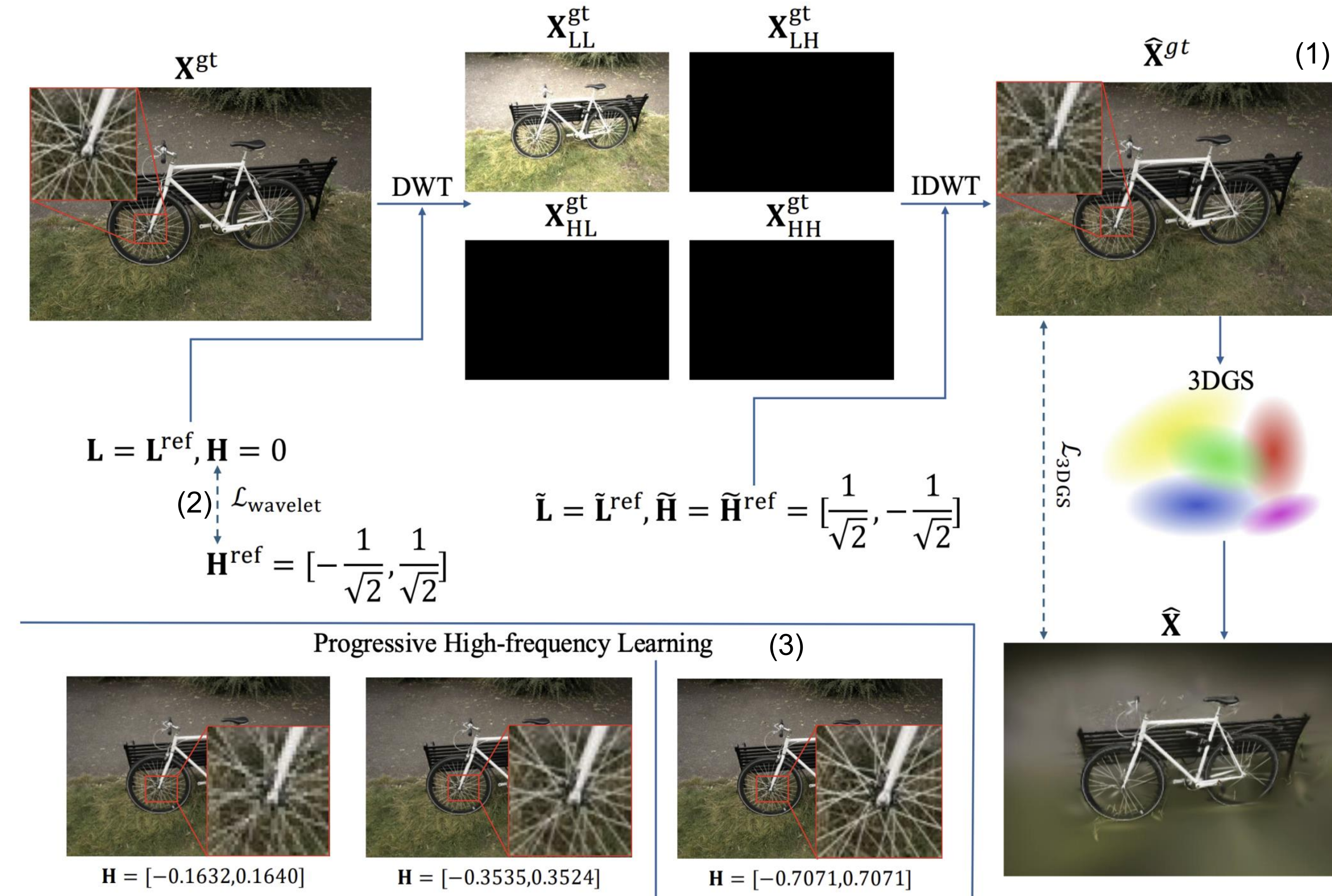
- The Inverse DWT reconstructs the image from the subbands:

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

where $\tilde{\mathbf{L}}$ and $\tilde{\mathbf{H}}$ are low-pass and high-pass *synthesis* matrices, constructed similarly to the analysis matrices above.



Methodology



Quantitative Results

	PSNR ($\uparrow$)	#G ($\downarrow$)
3DGS	20.40	272K
Opti3DGS (*)	19.59	247K
AutoOpti3DGS	20.39	224K

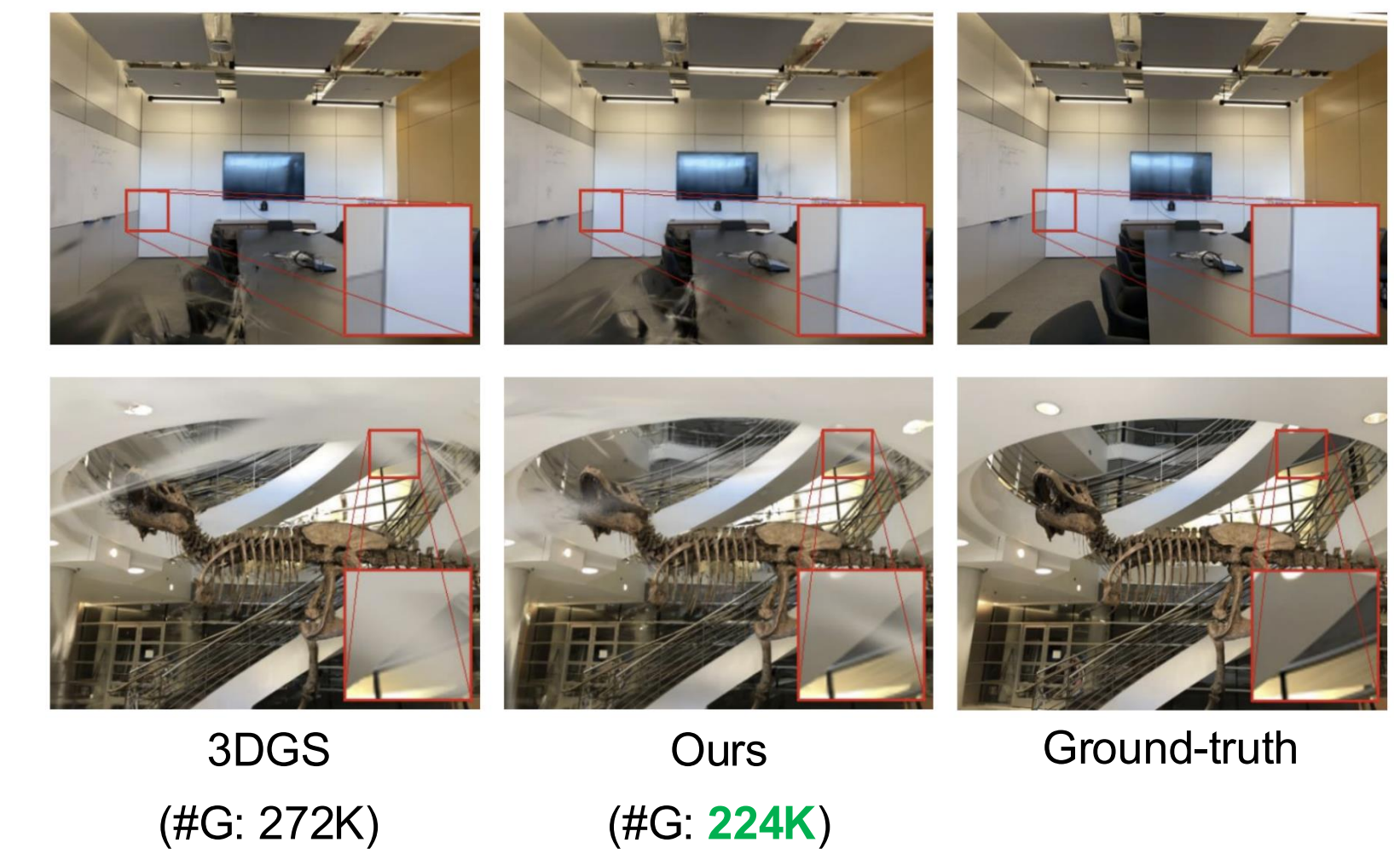
3-view LLFF dataset

(*) a predecessor to our method, whose modulation schedule is based on mean filtering with pre-defined, gradually decreasing kernel sizes

	PSNR ($\uparrow$)	#G ($\downarrow$)
3DGS	19.30	701K
Opti3DGS	19.19	636K
AutoOpti3DGS	19.24	566K

12-view Mip-NeRF 360 dataset

Qualitative Results



3-view LLFF dataset

- AutoOpti3DGS initializes $\mathbf{H} = \mathbf{0}$, restricting 3DGS to reconstructing coarse structures (1). An auxiliary loss, $\mathcal{L}_{wavelet}$, encourages $\mathbf{H}$ to converge towards the Haar references (2). As $\mathbf{H}$ converges, the input images are automatically injected with more fine-grained frequencies (3), without having to define a modulation schedule.
- The coarse-to-fine modulation avoids premature formations of fine Gaussians, only reconstructing them when necessary, thus optimizing Gaussian counts.