



WaveletGaussian: Wavelet-domain Diffusion for Sparse-view 3D Gaussian Object Reconstruction

Hung Nguyen, Runfa Li, An Le, Truong Nguyen
Video Processing Lab, University of California San Diego



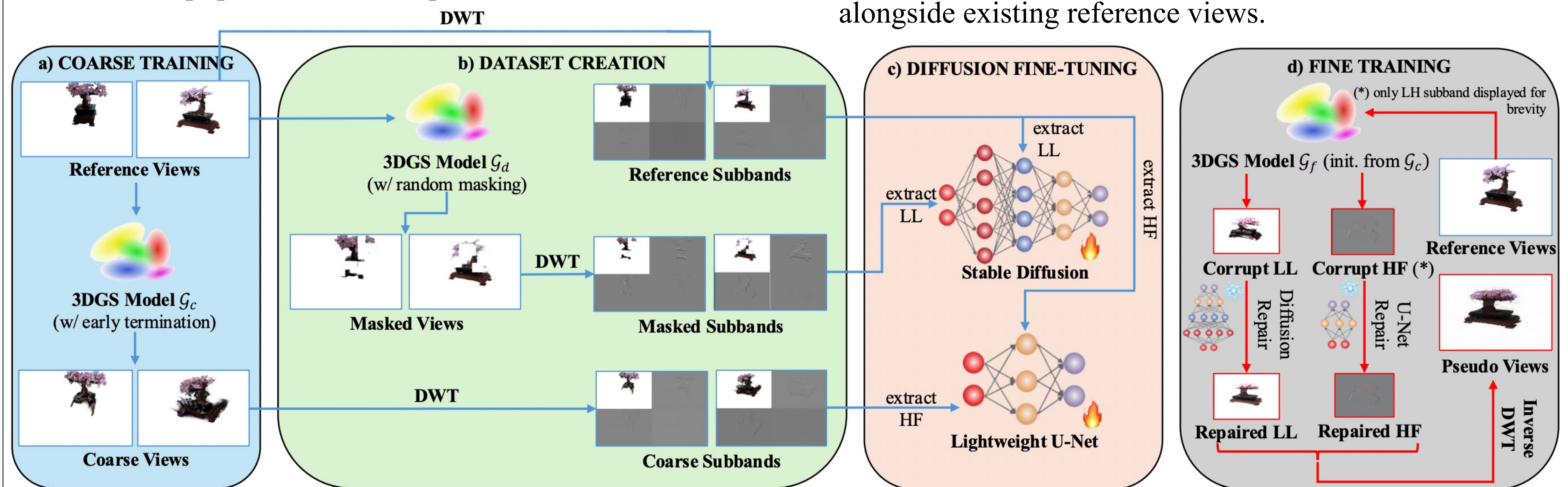
Motivation

- 3D Gaussian Splatting (3DGS) typically relies on dense training views. With sparse views, the rendering quality is severely compromised because the reconstructed geometry is under-constrained.
- Prior work leverages diffusion models to enhance the corrupted renders, which are then used as pseudo ground-truths to emulate dense-view training. However, this incurs significant computation due to fine-tuning the diffusion model to scene-specific data.
- We introduce **WaveletGaussian**, an efficient 3DGS framework for object reconstruction, leveraging diffusion-based repair in low-resolution wavelet space to rectify sparse-view artifacts.

Methodology

- Coarse Training:** a coarse 3DGS is trained to capture the overall geometry.
- Dataset Creation:** a masked 3DGS is trained to generate clean-corrupt pairs in wavelet space.

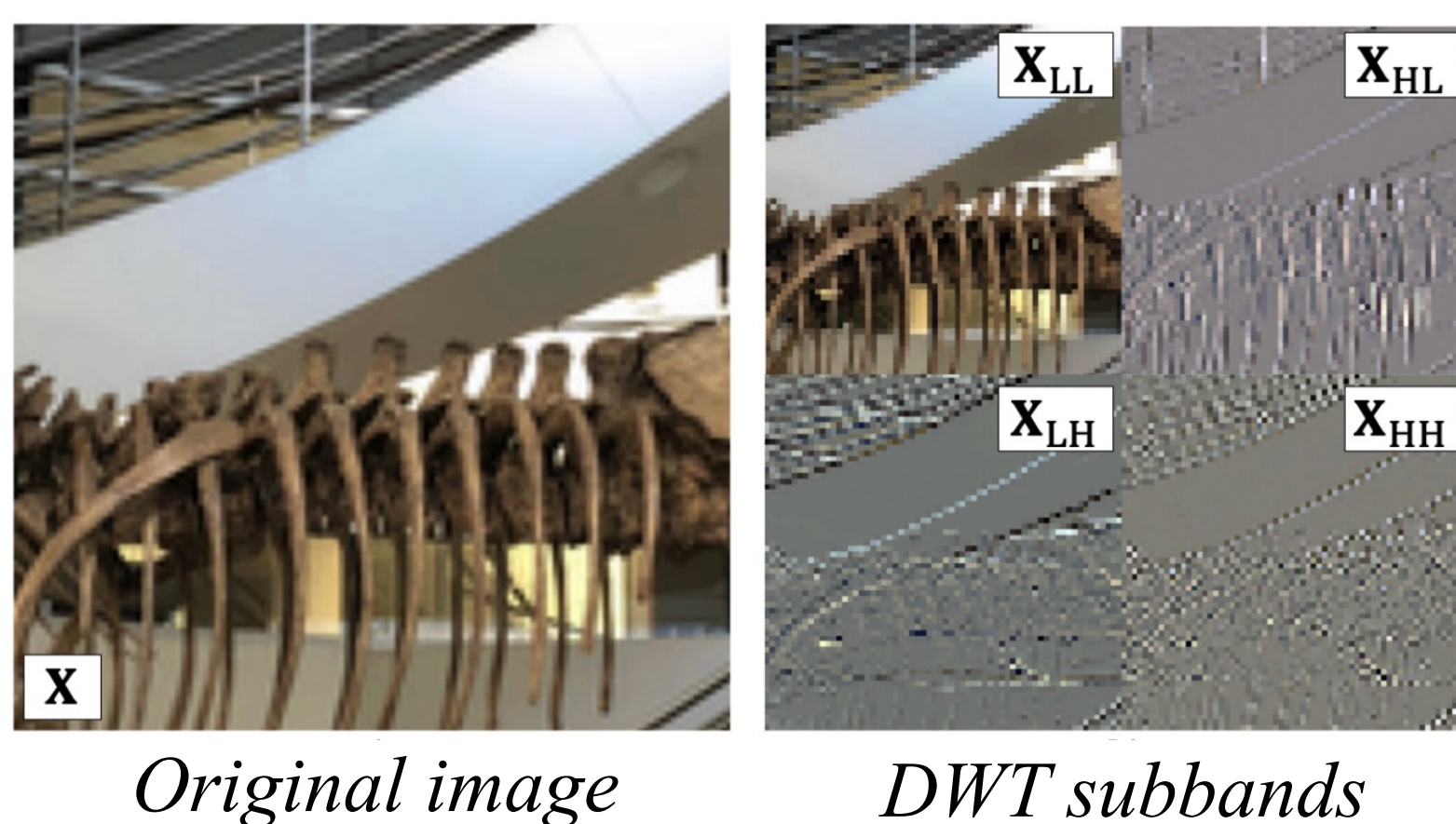
- Diffusion Fine-tuning:** a diffusion model learns to repair the LL subband while a lightweight U-Net handles high-frequency (HF) ones.
- Fine Training:** The coarse 3DGS is refined using pseudo views reconstructed from the repaired subbands, alongside existing reference views.



Discrete Wavelet Transform (DWT)

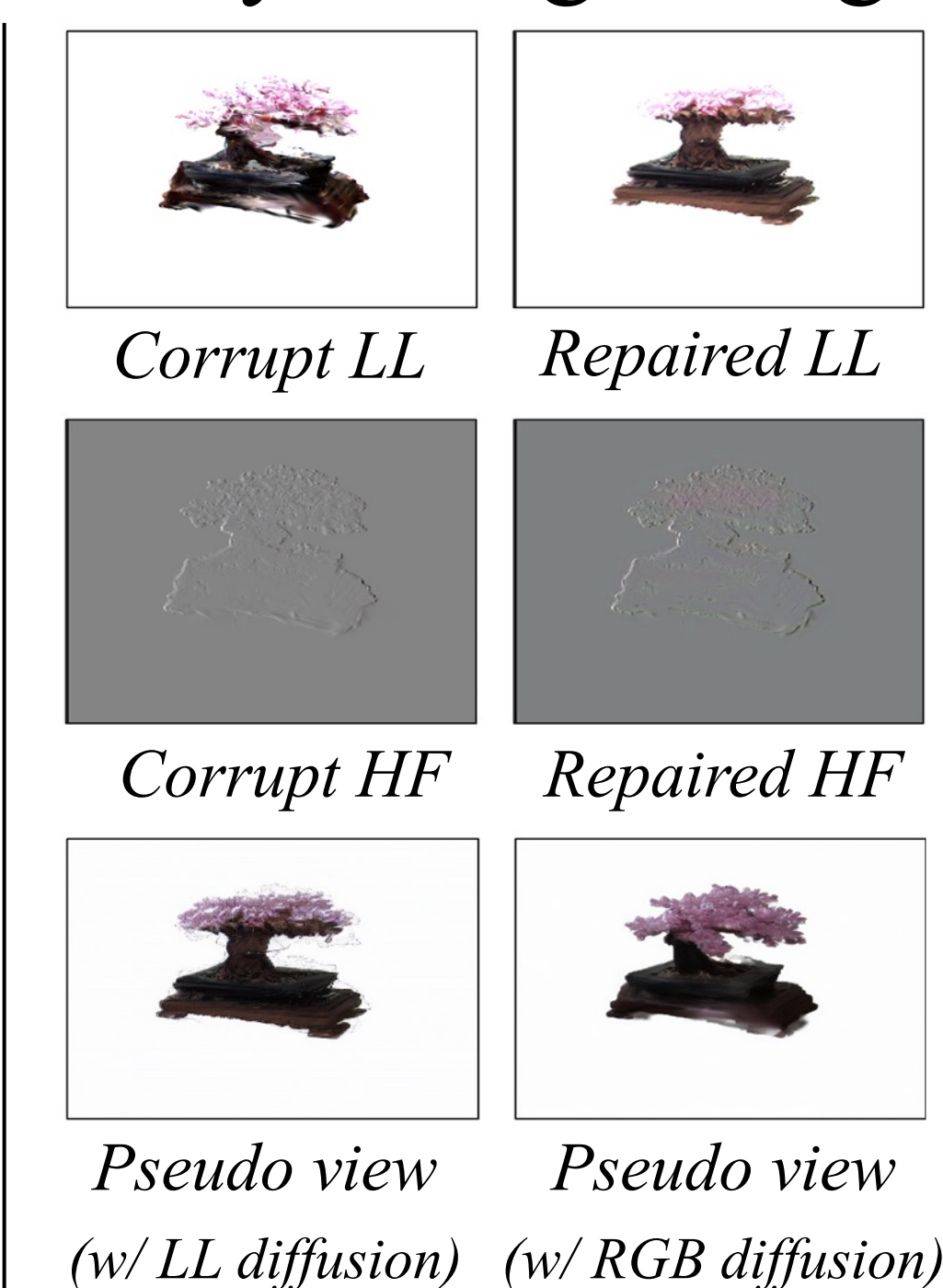
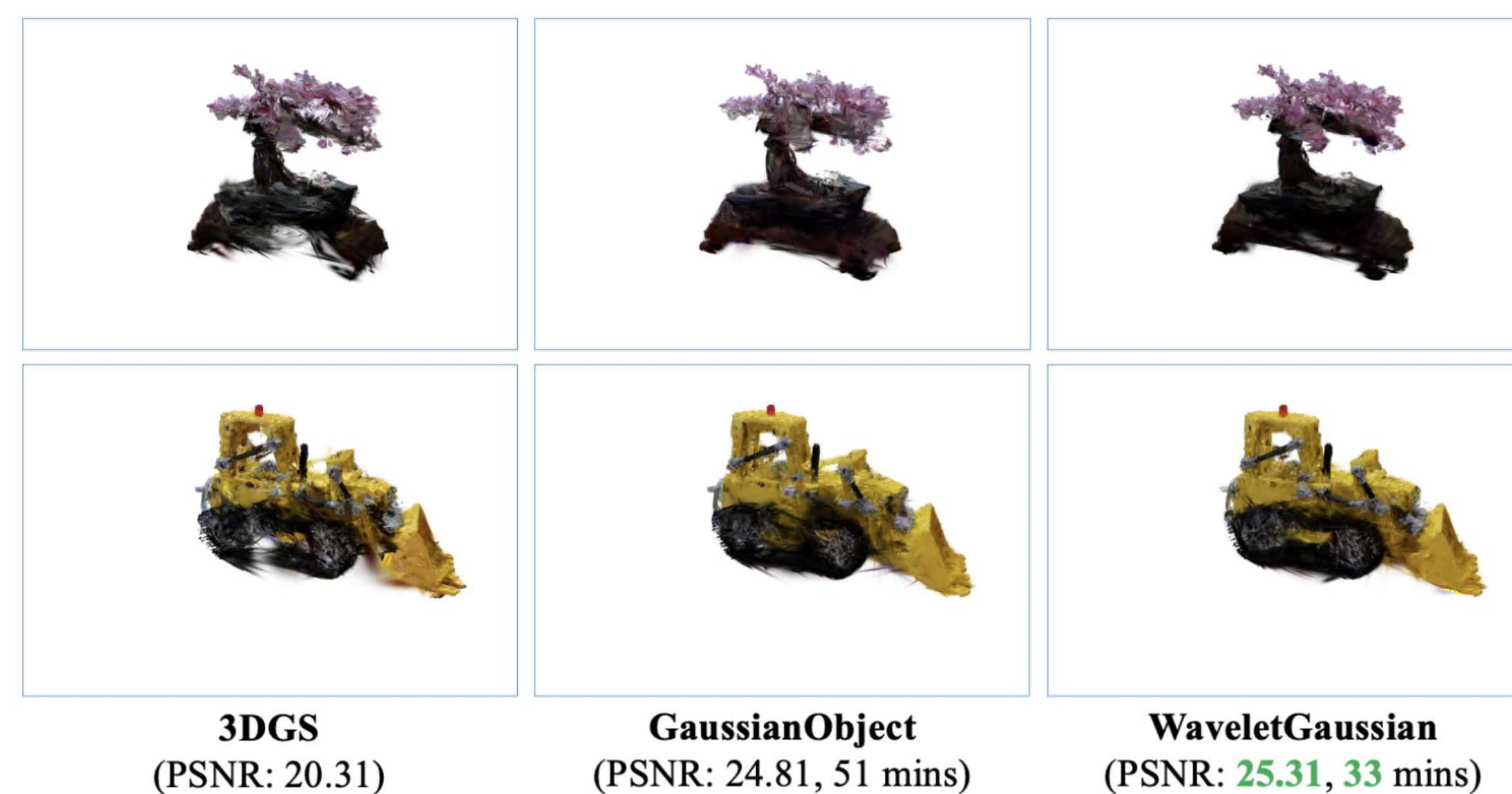
- The Forward DWT decomposes an input image $\mathbf{X}$ into four lower-resolution subbands, using analysis matrices $\mathbf{L}$ and $\mathbf{H}$:

$$\mathbf{X}_{LL} = \mathbf{LXL}^T, \mathbf{X}_{LH} = \mathbf{HXL}^T, \mathbf{X}_{HL} = \mathbf{LXH}^T, \mathbf{X}_{HH} = \mathbf{HXH}^T$$
- Using the Haar wavelet, $\mathbf{L}$ and $\mathbf{H}$ are constructed by shifting the 1D low-pass filter $\ell = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$ and high-pass filter $h = [-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$ along rows.
- The $\mathbf{X}_{LL}$, $\mathbf{X}_{LH}$, $\mathbf{X}_{HL}$, $\mathbf{X}_{HH}$ subbands emphasize the coarse structures, horizontal, vertical and diagonal details.
- From the subbands, the Inverse DWT reconstructs the original image using synthesis matrices $\tilde{\mathbf{L}}$ and $\tilde{\mathbf{H}}$.



Qualitative Results

- Our framework achieves comparable results to the RGB-diffusion baseline, GaussianObject [1], while requiring considerably less training times.
- The diffusion model repairs in lower-resolution LL space to save computation, but the HF information is still refined by the lightweight U-Net and retained via the Inverse DWT.



Quantitative Results

Method	PSNR	SSIM	LPIPS	Time
3DGS	20.31	0.899	0.108	-
GaussianObject	24.81	0.935	0.050	51 mins
WaveletGaussian (Ours)	25.31	0.939	0.047	33 mins

4-view Mip-NeRF 360 [2] dataset

Masking	LL repair	HF repair	PSNR	Time
✗	✗	✗	24.81	51 mins
✓	✗	✗	25.10	43 mins
✓	✓	✗	24.99	30 mins
✓	✓	✓	25.31	33 mins

Ablation studies

References

- Yang et al., "GaussianObject: High-Quality 3D Object Reconstruction from Four Views with Gaussian Splatting," SIGGRAPH Asia 2024.
- Barron et al., "Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields," CVPR 2022.