2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Frequency-Guided 3D Gaussian Splatting for Challenging Low-Light View Synthesis

Zhaoyuan Mai, Bi Zeng, Boquan Zhang†, Tianle Zeng, Jingxuan Lu, Jiarong Feng

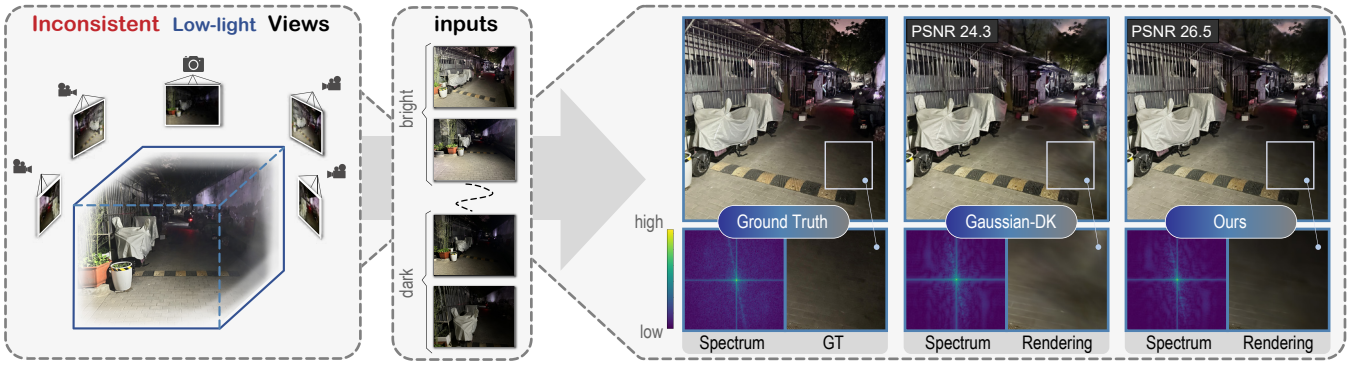

Fig. 1. Exposure-inconsistent inputs cause unnatural appearance in existing methods. We achieve better spatial reconstruction while maintaining spectral characteristics that balance noise suppression and detail preservation.

***Abstract*— Robust 3D scene understanding is crucial for autonomous robots, but degrades sharply in low-light environments where sensor noise and illumination inconsistencies corrupt visual inputs. Even 3D Gaussian Splatting (3DGS), while efficient for real-time reconstruction, produces unstable and artifact-prone results under such conditions, limiting its reliability for navigation and mapping. To address these challenges, we propose a 3DGS-based framework for reconstructing clear scenes under low-light conditions. Firstly, we employ a frequency-aware modulator that operates on spectral components to decouple and suppress sensor noise from structural signals, providing a clean input for reconstruction. To refine the 3D model and ensure its compactness for onboard deployment, we introduce an adaptive denoising mask guided by dynamically updated statistics of rendering contribution and stability, which filters transient artifacts caused by sensor noise. Finally, a multi-view frequency consistency constraint is enforced to ensure the global coherence of the reconstructed model's appearance, which is critical for consistent mapping. Experiments on challenging low-light datasets demonstrate that our method achieves state-of-the-art reconstruction quality while significantly reducing model storage by approximately 46.4% and maintaining real-time rendering speeds.**

I. INTRODUCTION

Real-time, high-fidelity 3D scene reconstruction is an important capability in modern robotics, supporting tasks such as autonomous navigation, mapping, object manipulation, and human-robot interaction. The recent emergence of 3D Gaussian Splatting (3DGS) [1] has marked a significant leap forward, offering explicit, photorealistic scene representations that can be rendered in real time. This combination of speed and quality makes it a highly promising paradigm for building the persistent and detailed world models required by autonomous systems. However, deploying such systems in low-light environments with inconsistent exposures remains challenging, as current methods struggle to maintain multi-view geometric consistency under adverse conditions.

Low-light environments impair robotic perception through high sensor noise, unpredictable exposure variations, and localized artifacts from active light sources [2]. These factors violate the multi-view geometric consistency assumption central to 3D reconstruction, leading to unstable models with erroneous geometry and flickering textures. Such inconsistencies impair downstream tasks such as Simultaneous Localization and Mapping (SLAM) and navigation.

Neural rendering methods addressing illumination and exposure variations can be divided into two groups. Implicit approaches, such as NeRF-W [3] and HDR-NeRF [4], model appearance changes via embeddings or exposure conditioning, but their high computational costs make them impractical for real-time robotic applications. Explicit approaches, exemplified by Gaussian-DK [2], integrate lightness mapping into the 3DGS pipeline to handle exposure inconsistencies while preserving real-time efficiency.

Despite these advances, current methods still struggle with severe illumination variations and sensor noise. As shown in Figure 1, input views with significant exposure variation often lead to reconstruction artifacts such as inconsistent illumination and unnatural appearance. A closer analysis reveals systematic deficiencies: (1) Gaussian-DK exhibits spatial illumination inconsistencies that manifest as low-frequency spectral mismatches across views, causing flick-

This paper is part of the outcomes from the Cultivation Project for Teaching Achievements in Undergraduate Higher Education at Guangdong University of Technology: "Student-Centered Immersive Artificial Intelligence Curriculum Development and Practice" (Project No. 263116029; Document No. GDUT-Ed-2023-78).

Authors are with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou {maizhaoyuan01, 1010466868, 1043146492, 2542184285}@qq.com, {zb9215, ivancheung}@gdut.edu.cn
†Corresponding author

ering artifacts. (2) The reconstruction shows altered high-frequency characteristics compared to ground truth, reflecting the challenge of balancing noise suppression with texture preservation.

Building upon 3DGS, we propose a frequency-guided framework that utilizes spectral properties to address the fundamental problem of multi-view exposure inconsistencies in low-light 3D reconstruction. Firstly, we employ a Frequency-aware Lightness Modulator (FLM) operating in the feature bottleneck to independently process magnitude and phase spectra, effectively suppressing noise while preserving structural details. Second, we incorporate an Adaptive Denoising Mask that learns to filter unstable Gaussians based on long-term rendering statistics, unifying noise removal with model compression. Finally, we enforce multi-view frequency consistency to maintain coherence among low-frequency components across neighboring viewpoints, ensuring globally consistent appearance. In summary, the major contributions of our work are as follows:

- We propose a frequency-domain framework that enhances robustness through FLM for spectral modulation, and a multi-view frequency constraint for global consistency.
- We introduce an adaptive denoising Mask that learns probabilistic retention of Gaussians, unifying noise filtering and model sparsification.
- Experiments on Gaussian-DK dataset show state-of-the-art quality and 46.4% model compression with real-time performance.

II. Related Work

A. Novel View Synthesis

Novel view synthesis aims to generate photorealistic images from a set of existing images or viewpoints. Traditional methods utilize light fields [5], multi-view stereo [6], and image-based rendering [7]. Neural Radiance Fields (NeRF) [8] revolutionized this field through implicit volumetric representation, with subsequent works improving quality [9], efficiency [10], and generalization [11]. 3D Gaussian Splatting (3DGS) [1] addresses this limitation using explicit anisotropic Gaussians with differentiable rasterization. Recent extensions cover dynamic scenes [12], anti-aliasing [13], large-scale reconstruction [14], and efficiency optimization [15], [16]. Nevertheless, most methods assume consistent illumination. While appearance embeddings [3] and multi-exposure modeling [4] provide partial solutions, severe illumination inconsistencies in low-light environments remain challenging.

B. Low-Light Image Processing

Low-light image processing addresses the challenges of limited photon counts, high sensor noise, and reduced dynamic range. Traditional methods include Retinex-based decomposition [17], histogram equalization [18], and camera response modeling [19]. Deep learning approaches range from direct enhancement networks [20], [21] to physics-inspired models [22], with recent self-supervised [23] and zero-shot methods [24] reducing paired data requirements. For 3D reconstruction, several methods have emerged to enhance low-light scene reconstruction [25], [26], [27]. Gaussian-DK [2] addresses multi-view inconsistencies through exposure-aware rendering and learnable lightness mapping. However, these methods struggle to differentiate genuine illumination variations from artifacts caused by sensor noise and imaging conditions, especially when ensuring consistency across views with varying exposures and noise profiles. To address this limitation, frequency-domain perspectives offer a promising alternative.

C. Frequency-Domain Methods in Computer Vision

Frequency-domain analysis provides powerful tools for understanding and manipulating visual content. Classical signal processing theory [28] established that image structure is predominantly encoded in the phase spectrum, while magnitude governs energy distribution and contrast, enabling principled separation of content and artifacts. Traditional methods leverage this through enhancement [29], homomorphic filtering [30], and denoising [31]. Deep learning incorporates frequency processing via Fourier losses [32] and frequency-aware architectures [33]. In neural rendering, frequency-domain methods have shown promise: WaveNeRF [34] introduces wavelet frequency decomposition for generalizable NeRF, and FreGS [35] applies frequency regularization to improve 3D Gaussian densification by minimizing spectral discrepancies between rendered and ground truth images. However, prior works mainly target single-view enhancement or geometric accuracy, while frequency-domain constraints for cross-view illumination consistency remain largely unexplored, motivating our approach.

III. Preliminaries

3D Gaussian Splatting (3DGS). 3DGS [1] represents scenes using K anisotropic 3D Gaussian primitives that preserve differentiability while enabling real-time rendering. Each Gaussian G_i is parameterized by position $\boldsymbol{\mu}_i \in \mathbb{R}^3$, covariance matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{3\times3}$, opacity $\alpha_i \in [0,1]$, and view-dependent color c_i using spherical harmonics coefficients. The covariance is decomposed as $\boldsymbol{\Sigma}_i = \mathbf{R}_i\mathbf{S}_i\mathbf{S}_i^\top\mathbf{R}_i^\top$ where $\mathbf{R}_i$ and $\mathbf{S}_i$ are rotation and scaling matrices. Rendering employs tile-based differentiable rasterization via alpha-compositing:

$$C(\mathbf{x}') = \sum_{i=1}^{N} c_i\alpha_i' T_i, \quad T_i = \prod_{j<i}(1-\alpha_j'), \tag{1}$$

where $\alpha_i' = \alpha_i G_i'(\mathbf{x}')$ is the projected opacity of Gaussian i.

Gaussian-DK Architecture. We build our method upon Gaussian-DK [2], which extends each 3D Gaussian with learnable light features $\mathbf{f}_i \in \mathbb{R}^8$ that encode local illumination properties. The framework incorporates exposure-level conditioning through:

$$\Delta E = \frac{K \times T \times ISO}{A^2}, \tag{2}$$

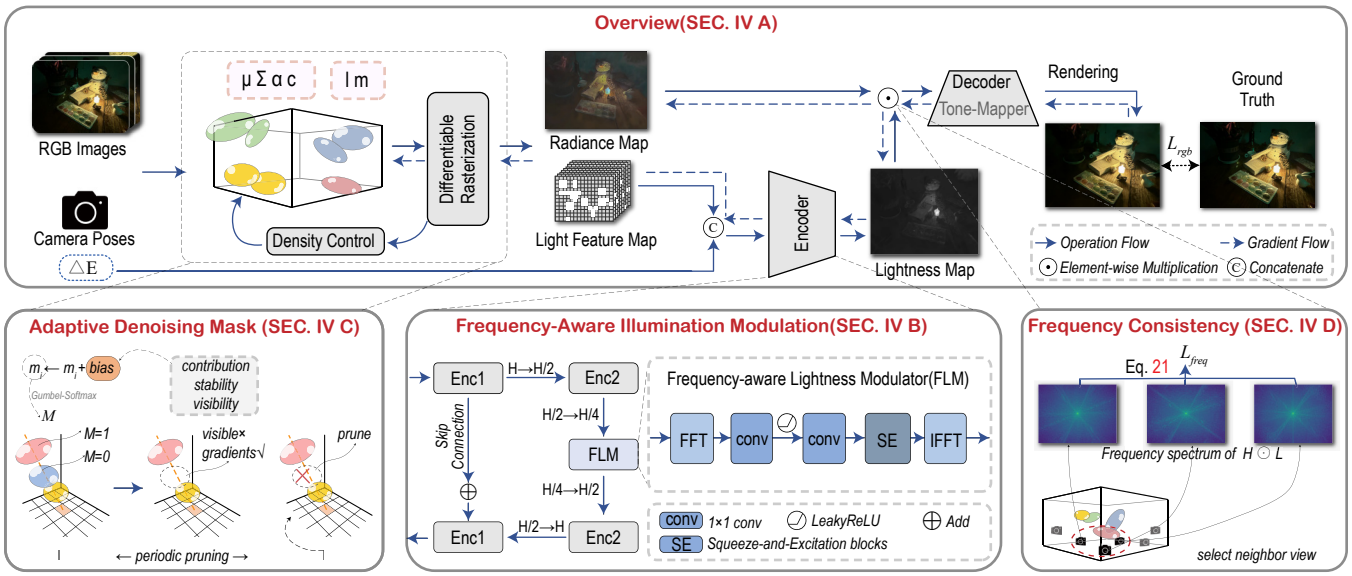

Fig. 2. **Overview of our framework.** Built upon explicit 3D scene representations using Gaussian primitives, our method addresses the fundamental challenges of low-light novel view synthesis through frequency-domain processing and adaptive primitive filtering. The Frequency-aware Lightness Modulator processes magnitude and phase spectra independently at the LightnessMapper bottleneck, the Adaptive Denoising Mask filters Gaussian primitives based on long-term rendering statistics, and Multi-view Frequency Consistency enforces illumination coherence across neighboring views. These components work synergistically within the rendering pipeline to achieve robust low-light novel view synthesis with significant storage reduction.

where T, A, and ISO represent exposure time, aperture, and sensor sensitivity respectively.

During rendering, 3D Gaussians are rasterized to produce a radiance map H and light feature map $\mathbf{F}$. The final pixel values are computed as:

$$Z = \mathcal{T}(L \odot H), \tag{3}$$

where L is the lightness map generated by a LightnessMapper network from $\mathbf{F}$ and ΔE, and $\mathcal{T}$ denotes the ToneMapper network.

IV. Method

A. Overview

Figure 2 illustrates the overview of our pipeline. Starting from Structure from Motion (SfM) initialization, we observe that low-light scenes exhibit distinct frequency characteristics where structural information and artifacts manifest differently across spectral components. Building upon this insight, our approach integrates frequency-aware processing throughout the rendering pipeline. The Frequency-aware Lightness Modulator operates within the LightnessMapper network, decomposing features into frequency components and processing magnitude and phase spectra independently to naturally separate structural information from noise artifacts. Then, we introduce an Adaptive Denoising Mask mechanism that addresses the challenge of distinguishing stable geometry from transient artifacts through learnable retention probabilities. Additionally, our Multi-view Frequency Consistency constraint enforces illumination coherence across neighboring viewpoints by operating in the frequency domain, focusing on low-frequency stability while preserving view-dependent details.

B. Frequency-Aware Illumination Modulation

Low-light 3D scene reconstruction is particularly vulnerable to sensor noise and localized halos, which are often introduced by high ISO amplification and active light sources. In the spatial domain, these degradations are tightly entangled with structural signals at the pixel level, making illumination adjustment challenging without simultaneously damaging fine details. In contrast, the frequency domain provides a natural separation of these components. According to Fourier analysis [28], structural information is mainly encoded in the phase spectrum, while the magnitude spectrum captures the overall energy distribution and contrast. This separation allows us to treat structure and degradation differently, rather than being forced to handle them jointly in the spatial domain. Specifically, nighttime images typically exhibit halos from bright light sources and noise from low-intensity regions. These distortions manifest in different frequency bands: halos often lead to the loss of high-frequency signals, while sensor noise appears as high-frequency interference spread across the spectrum. Consequently, such artifacts possess inconsistent frequency characteristics, suggesting that frequency-domain processing provides a principled means to disentangle and suppress them.

Building on this insight, we propose the Frequency-aware Lightness Modulator (FLM). The core idea is to decompose illumination features into independent magnitude and phase components, so that noise and halo artifacts can be selectively suppressed while preserving structural details.

Given bottleneck features $\mathbf{x} \in \mathbb{R}^{C\times H\times W}$ from the Gaussian-DK illumination module, FLM first applies a 2D FFT:

$$\mathcal{X} = \mathcal{F}(\mathbf{x}), \quad \mathcal{A} = |\mathcal{X}|, \quad \mathcal{P} = \angle\mathcal{X}, \tag{4}$$

where $\mathcal{F}$ denotes the 2D Fourier transform, $\mathcal{A}$ and $\mathcal{P}$ are the magnitude and phase spectra.

Magnitude and phase are enhanced independently via channel-attentive networks:

$$\mathcal{A}' = \mathcal{A} + g_A(w_A \odot \sigma(g_A(\mathcal{A}))), \tag{5}$$

$$\mathcal{P}' = \mathcal{P} + g_P(w_P \odot \sigma(g_P(\mathcal{P}))), \tag{6}$$

where g_A and g_P denote 1×1 convolutional layers, w_A and w_P are channel attention weights dynamically computed by squeeze-and-excitation [36] blocks, $\odot$ represents element-wise multiplication, and $\sigma(\cdot)$ denotes the LeakyReLU activation function.

The enhanced spectra are recombined via inverse FFT:

$$\mathbf{x}' = \mathcal{F}^{-1}(\mathcal{A}' \cdot e^{i\mathcal{P}'}), \tag{7}$$

Where $\mathcal{F}^{-1}$ is the inverse Fourier transform.

To reduce FFT computation cost and leverage the rich global illumination context available at low resolution, FLM is placed at the bottleneck of the LightnessMapper at $H/4 \times W/4$ resolution, which ensures both efficiency and globally consistent feature enhancement. Operating at the deepest U-Net layer ensures that enhanced features are globally consistent and readily decoded to full resolution, providing a solid foundation for subsequent multi-view consistency enforcement.

C. Adaptive Denoising Mask

Original 3DGS achieves effective point cloud densification through cloning and splitting while filtering low-contribution points. Although demonstrating efficiency in point management, this mechanism struggles in low-light scenarios where sensor noise and inter-view illumination inconsistencies cause Gaussian primitives to exhibit heterogeneous behaviors. While structural geometry maintains stable rendering contributions, transient artifacts generate spurious contributions from pseudo-Gaussians, degrading reconstruction quality.

For each Gaussian, we maintain learnable mask scores $m_i = [m_i^{\text{keep}}, m_i^{\text{drop}}]$ that determine the probability of retention. Binary masks $M_i \in \{0, 1\}$ are sampled via Gumbel-Softmax from these scores and integrated within the rasterization process:

$$c(x) = \sum_{i=1}^{N} M_i \cdot c_i \cdot \alpha_i \cdot T_i, \tag{8}$$

$$T_{i+1} = M_i \cdot (1 - \alpha_i) \cdot T_i + (1 - M_i) \cdot T_i, \tag{9}$$

When $M_i = 1$, the Gaussian contributes normally to color accumulation and consumes transmittance based on its α_i. When $M_i = 0$, the Gaussian's color contribution is masked out and its transmittance consumption is skipped, ensuring correct forward pass computation. Crucially, masked Gaussians still participate in the forward computation and receive meaningful gradients for dynamic probability adjustment:

$$\frac{\partial \mathcal{L}}{\partial M_i} = \alpha_i \cdot T_i \cdot \frac{\partial \mathcal{L}}{\partial c(x)} \cdot (c_i - b_{i+1}), \tag{10}$$

where $b_{i+1} = \sum_{j=i+1}^{N} M_j \cdot c_j \cdot \alpha_j \cdot T_j$ represents the cumulative contribution from subsequent Gaussians.

However, to further refine this learning process, especially for Gaussians whose retention probabilities are uncertain, we introduce a contribution-stability prior to bias the optimization. This prior is derived from three intrinsic statistics tracked for each Gaussian i at iteration t:

$$w_i^{(t)} = \frac{1}{|\mathcal{P}_i^{(t)}|} \sum_{p \in \mathcal{P}_i^{(t)}} (\alpha_{i,p} \cdot T_{i,p}), \tag{11}$$

$$v_i^{(t)} = \mathbf{1}[\mathcal{P}_i^{(t)} > 0], \tag{12}$$

$$s_i^{(t)} = |w_i^{(t)} - c_i^{(t-1)}|, \tag{13}$$

where $\mathcal{P}_i^t$ denotes the set of views where Gaussian i contributes to rendering. These statistics are maintained via EMA updates:

$$c_i^{(t)} = (1 - \rho)c_i^{(t-1)} + \rho w_i^{(t)}, \quad \rho = 0.1 \tag{14}$$

with similar updates for $v_i^{(t)}$ and $s_i^{(t)}$.

Based on min-max normalized statistics $\hat{c}_i, \hat{v}_i, \hat{s}_i \in [0, 1]$, we compute a prior retention probability:

$$r_i = \sigma(a \cdot \hat{c}_i + b \cdot \hat{v}_i - d \cdot \hat{s}_i), \tag{15}$$

where a, b, d are learnable global parameters that weigh the relative importance of the statistics, initialized to 1.0 and optimized jointly with the rest of the model. This formulation provides a flexible and stable prior. This prior guides the mask probability learning by adjusting the retention score m_i:

$$m_i \leftarrow m_i + \beta(t) \cdot \left[\log\left(\frac{r_i}{1 - r_i}\right), -\log\left(\frac{r_i}{1 - r_i}\right)\right], \tag{16}$$

where $\beta(t)$ is a scheduling factor that anneals the guidance strength. We employ a delayed linear annealing schedule: $\beta(t)$ remains constant at an initial value for the first 70% of training, and then linearly decays to zero. This allows the prior to exert a strong, consistent influence during the main learning phase, which then gradually diminishes to allow for fine-tuning. This adjustment is selectively applied only to Gaussians with uncertain retention probabilities (e.g., between 0.3 and 0.7), focusing the guidance on where it is most needed. In addition to enabling the adaptive mask to remove more points, we incorporate the following constraints:

$$L_m = \left(\frac{1}{N} \sum_{i=1}^{N} M_i\right)^2. \tag{17}$$

Finally, to eliminate Gaussians with consistently low retention probabilities, we perform a pruning step. Each Gaussian's mask is sampled 10 times, and any Gaussian that is never selected for retention is permanently removed. This stochastic pruning is executed during each densification step, and every 1000 iterations after densification.

D. Multi-View Frequency Consistency Constraint

For scenes captured under varying lighting conditions, ensuring consistent illumination across novel views is crucial for photorealism. Directly enforcing pixel-level consistency between views is often counterproductive, as it is highly sensitive to parallax and geometric details, which can penalize correct view-dependent effects. To address this, we propose a multi-view frequency consistency constraint that operates on the illumination component in the frequency domain.

Our constraint is formulated based on two principles of cross-view stability. First, we operate only on the frequency magnitude, as the phase spectrum encodes spatial positions that are highly sensitive to inter-view parallax. Second, we primarily target the low-frequency bands, which represent stable global illumination, while allowing high-frequency details, which are often entangled with view-dependent geometry, to vary.

a) Signal Selection: We apply our consistency constraint on the enhanced linear radiance map $H \odot L$ in log space, which represents the final illumination-adjusted radiance before tone mapping. This signal captures both the scene's intrinsic radiance and the learned illumination adjustments, making it an ideal target for physical consistency enforcement while preserving the benefits of learned enhancement.

b) Neighbor Selection: To compute the consistency loss for a given view i, we first identify a set of suitable neighboring views N_i. A robust neighbor selection strategy is key to avoiding erroneous constraints between disparate views. We define a neighbor j as valid only if it satisfies both geometric and photometric similarity criteria:

Geometric Proximity: The camera centers c_i and c_j must be spatially close. We use a normalized distance threshold, $d(c_i, c_j) < \rho \cdot E_{\text{scene}}$, where E_{scene} is the diagonal extent of the scene's camera rig and ρ is a fixed ratio. This ensures the criterion adapts to the scale of different scenes.

Viewing Direction Similarity: The views must observe the scene from a similar direction. We enforce this by ensuring the angle between the vectors from the camera centers to the scene centroid q is below a threshold θ_{max}. This filters out cameras that are physically close but look in opposite directions (e.g., on either side of an object).

Photometric Similarity: Among geometrically valid candidates, we rank them by a score

$$S_{ij} = \frac{\exp(-|EV_i - EV_j|)}{1 + d(c_i, c_j)}, \tag{18}$$

which prioritizes photometric similarity measured by Exposure Value (EV) difference. We select all qualified neighbors that satisfy the criteria to form N_i.

c) Loss Formulation: Our frequency consistency loss, L_{freq}, encourages the magnitude of frequency components of the lightness maps to be similar between a view i and its neighbors $j \in N_i$. First, we decompose the 2D lightness map L into K frequency bands using 2D Fourier Transform and a set of concentric ring masks, yielding frequency components $\{F_k(L)\}_{k=0}^{K-1}$, where $k = 0$ corresponds to the lowest frequency band.

The total loss is a weighted sum of Mean Squared Error (MSE) between the frequency magnitudes of the view i and its neighbors:

$$L_{\text{freq}} = \frac{1}{|N_i|} \sum_{j \in N_i} w_{EV}(i,j) \sum_{k=0}^{K-1} w_k \, \| \, |F_k(L_i)| - |F_k(L_j)| \, \|_2^2, \tag{19}$$

where $|\cdot|$ denotes the magnitude of the complex numbers. This loss incorporates two key weighting terms.

First, the EV-aware weight w_{EV} modulates the influence of a neighbor based on photometric difference. Views with very different exposures are expected to have different illumination maps, so we assign a lower weight to their consistency term using a gentle exponential decay:

$$w_{EV}(i,j) = \exp(-|EV_i - EV_j|). \tag{20}$$

Second, the frequency band weight w_k materializes our core hypothesis that low-frequency components are more consistent across views. We thus down-weight higher frequency bands exponentially:

$$w_k = \lambda^k, \quad \lambda \in (0,1). \tag{21}$$

This focuses the constraint on global illumination effects while allowing high-frequency geometric details to vary.

V. EXPERIMENTS

A. Datasets and Metrics

We evaluate our method on a challenging low-light dataset containing 12 real-world scenes (5 indoor and 7 outdoor), which was proposed by Gaussian-DK. The metrics used for evaluation include Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). In addition, we conducted a comparison of rendering efficiency, including metrics such as frames per second (FPS) and storage usage.

B. Experiment Details

We implemented our framework using PyTorch. Throughout the training process, we executed a total of 15,000 iterations with resolution downsampling factor of 4. For the contribution-aware adaptive masking, we employed EMA smoothing coefficient $\rho = 0.1$ with statistics updated every 50 iterations. The contribution weights were set as $a = 1.0$, $b = 1.0$, $d = 1.0$. For the multi-view frequency consistency constraint, we used neighbor radius $\rho = 1.5$ times the scene extent. The frequency decomposition used $K = 4$ bands with exponential decay weights $\lambda = 0.2$.

Regarding the computation of loss between rendered images and real images, we adhered to the approach outlined in 3D-GS, utilizing $\mathcal{L}_{\text{D-SSIM}}$ and $\mathcal{L}_1$. We assigned weights to the loss functions $\mathcal{L}_{\text{mask}}$ and $\mathcal{L}_{\text{freq}}$, setting them at $5\text{e}{-4}$, and $3\text{e}{-2}$, respectively. Notably, $\mathcal{L}_{\text{freq}}$ contributed to training after a $4,000$-iteration warmup period with gradual weight increase, while $\mathcal{L}_{\text{mask}}$ was applied throughout the entire training process.

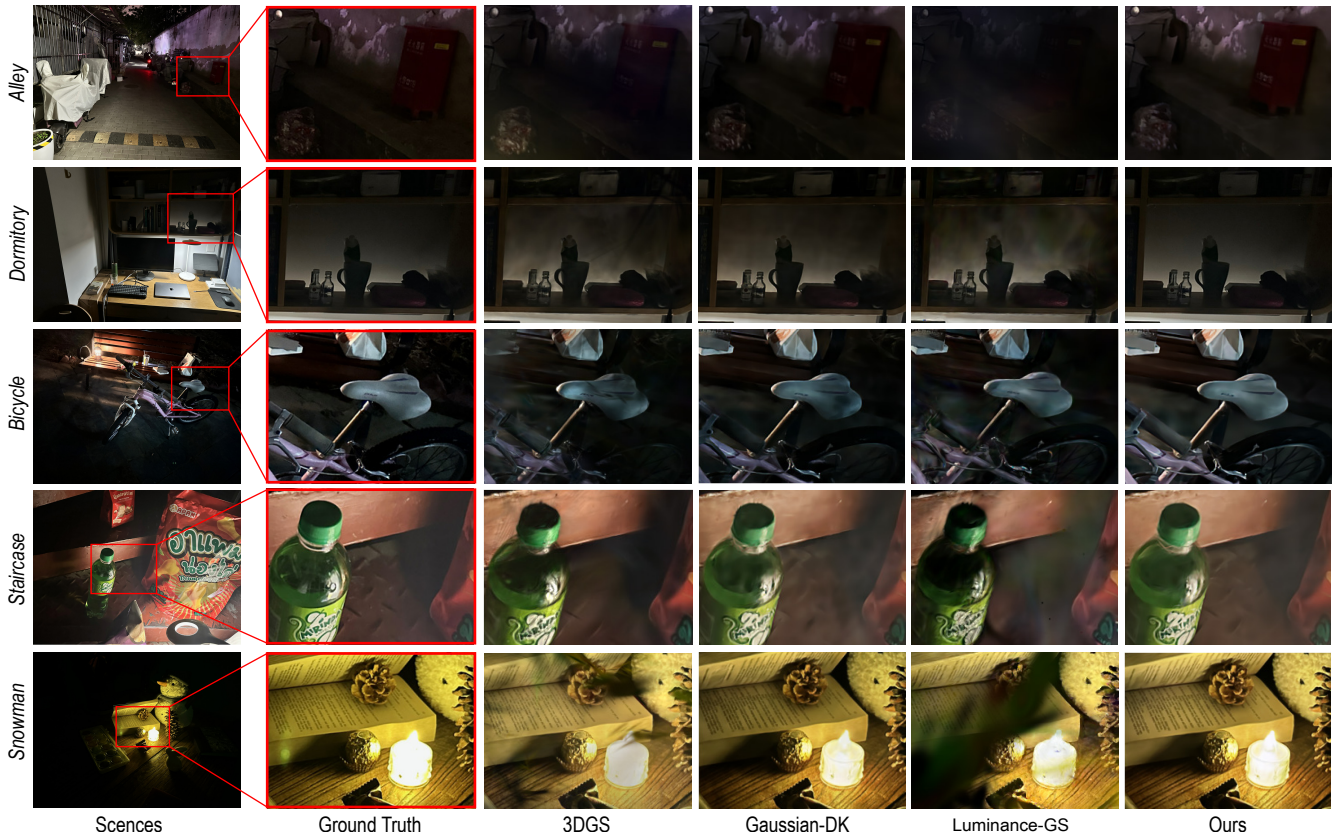

Fig. 3. **Qualitative comparisons of baselines and our method on Gaussian-DK dataset.** Our method achieves better noise suppression, cleaner shadow boundaries, and improved detail preservation. The highlighted regions show superior handling of complex illumination variations compared to baseline methods.

TABLE I

QUANTITATIVE COMPARISON ON GAUSSIAN-DK DATASETS. WE COLOR EACH CELL AS BEST, SECOND BEST, AND THIRD BEST.

Methods	Alley			Bicycle			Dormitory			Fence		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF-W	21.59	0.646	0.331	23.56	0.717	0.238	24.51	0.820	0.149	23.97	0.724	0.214
HDR-NeRF	17.87	0.558	0.488	19.95	0.608	0.267	21.23	0.684	0.279	19.66	0.578	0.372
3DGS	16.44	0.602	0.232	19.27	0.624	0.220	20.53	0.686	0.161	18.99	0.588	0.226
Gaussian-DK	21.91	0.773	0.140	24.80	0.809	0.131	25.64	0.857	0.087	24.94	0.801	0.131
Luminance-GS	16.66	0.686	0.225	19.97	0.657	0.223	20.45	0.707	0.193	19.50	0.717	0.239
Ours	23.01	0.803	0.128	25.65	0.835	0.128	26.70	0.880	0.077	25.13	0.809	0.127
Methods	**Flower**			**Kitchen**			**Livingroom**			**Piano**		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF-W	21.66	0.662	0.223	24.84	0.812	0.205	22.43	0.749	0.243	25.44	0.803	0.157
HDR-NeRF	18.54	0.517	0.359	22.27	0.757	0.338	20.77	0.651	0.385	21.52	0.682	0.240
3DGS	16.17	0.541	0.226	18.68	0.699	0.196	19.97	0.728	0.167	18.78	0.527	0.248
Gaussian-DK	21.50	0.728	0.135	25.54	0.844	0.117	24.41	0.826	0.125	27.31	0.833	0.106
Luminance-GS	17.76	0.692	0.275	16.43	0.730	0.251	19.76	0.781	0.201	21.35	0.675	0.224
Ours	21.91	0.732	0.138	26.50	0.856	0.115	25.46	0.842	0.114	28.25	0.852	0.104
Methods	**Street**			**Staircase**			**Snowman**			**Tree**		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF-W	22.85	0.638	0.293	23.96	0.809	0.115	–	–	–	21.66	0.649	0.260
HDR-NeRF	21.20	0.577	0.424	18.07	0.566	0.343	–	–	–	18.58	0.504	0.397
3DGS	17.04	0.542	0.271	18.30	0.583	0.198	19.47	0.630	0.230	19.87	0.597	0.195
Gaussian-DK	26.23	0.789	0.142	26.10	0.885	0.068	28.63	0.885	0.069	23.30	0.765	0.134
Luminance-GS	17.76	0.663	0.275	18.84	0.584	0.194	24.61	0.813	0.125	20.03	0.615	0.193
Ours	26.63	0.796	0.145	27.05	0.893	0.067	29.16	0.882	0.068	24.35	0.800	0.126

For optimization, we employed the Adam optimizer, with distinct learning rates assigned to different components: the 3D Gaussians retained a consistent learning rate as per the official implementation, the ToneMapper and LightnessMapper networks used learning rates of $1e{-3}$, while the adaptive mask learning rate was set to $1e{-2}$. All experiments were

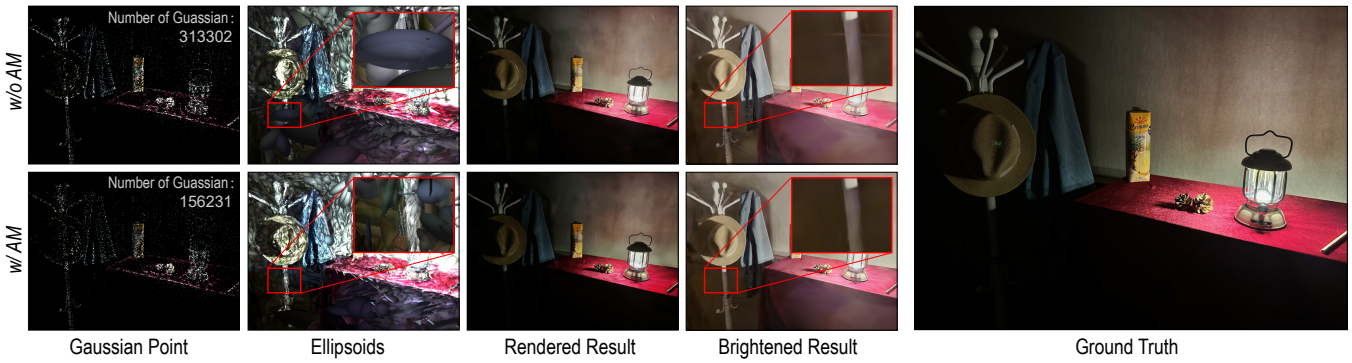

Fig. 4. Comparison of images with and without the use of adaptive denoising mask. We visualize Gaussian center points, ellipsoids, and rendered results using Piano scene. The brightened result shows enhanced illumination by modulating exposure levels through camera response modeling.

performed on an NVIDIA RTX 2080Ti GPU.

We compare our method with NeRF-W [3], HDR-NeRF [4], 3DGS [1], Gaussian-DK [2], and Luminance-GS [26]. For Luminance-GS, we adopt the official variance configuration, which is designed for varying-exposure multi-view data.

C. Experimental Results

Quantitative Analysis. The quantitative results are presented in Table I. Our method achieves the best performance across the 12 challenging low-light scenes from the dataset introduced by Gaussian-DK [2]. It consistently outperforms all baseline methods in key metrics. Notably, NeRF-based approaches such as NeRF-W and HDR-NeRF, output completely black renderings on the "Snowman" scene. Compared to the strongest baseline, Gaussian-DK, our method shows a clear average improvement of 0.79 dB in PSNR and 0.016 in SSIM, as well as an average reduction of 0.003 in LPIPS. These results confirm the superior accuracy and perceptual quality of our approach.

Qualitative Analysis. Figure 3 provides a qualitative comparison, showing that our method produces visually superior results. Our approach is particularly effective at handling noise and complex lighting, which are common challenges in dark scenes. For instance, in the Alley scene, our method successfully reconstructs fine details in the shadowed wall that are lost in other methods, demonstrating its ability to preserve texture and structure amidst sensor noise. Similarly, in the Dormitory scene, the shadow transitions behind the objects are noticeably smoother. This highlights our method's strength in producing more realistic and coherent lighting.

D. Ablation Study

To validate the effectiveness of our proposed components, we conduct ablation experiments on the Gaussian-DK dataset. Table II presents the progressive improvement as we incrementally add each component to the baseline Gaussian-DK model.

Component Analysis. The results demonstrate that each component provides distinct and complementary benefits. The Adaptive Denoising Mask (AM) significantly reduces storage overhead by approximately 46.4% while maintaining

TABLE II

ABLATION STUDY ON OUR PROPOSED COMPONENTS.

AM	FL	FC	PSNR ↑	SSIM ↑	LPIPS ↓	FPS ↑	Storage(MB) ↓
×	×	×	25.03	0.816	0.115	49	153
✓	×	×	25.09	0.818	0.115	58	82
✓	✓	×	25.68	0.828	0.113	49	82
✓	✓	✓	**25.82**	**0.832**	**0.112**	**49**	**82**

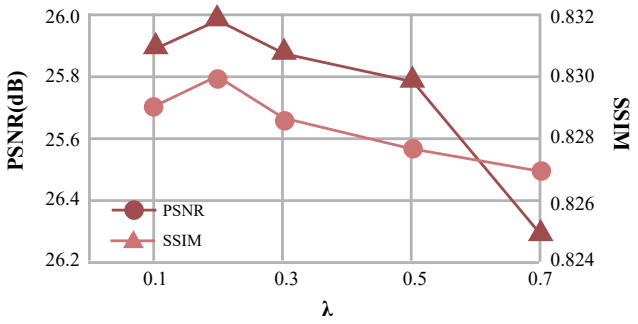

Fig. 5. Performance curve of frequency band weight λ sensitivity analysis. Experiments are conducted on the Gaussian-DK dataset. We vary λ for different frequency weighting strategies.

rendering quality, demonstrating its effectiveness in filtering unstable Gaussians. As shown in Figure 4, the mask successfully identifies and removes redundant Gaussians while preserving essential scene geometry, with the lightup visualization revealing details in shadow regions through modulated exposure levels. The Frequency-aware Lightness Modulator (FL) serves as the foundation for noise suppression and detail preservation, achieving a substantial improvement in PSNR and SSIM. The Frequency Consistency constraint (FC) enhances cross-view illumination coherence. The combination of all three components achieves optimal performance across all metrics, validating our design choices.

Frequency Band Weight Analysis. We analyze the impact of the frequency decay weight λ on rendering quality. As shown in Figure 5, different λ values exhibit distinct performance characteristics. A smaller λ heavily penalizes high-frequency components, leading to over-smoothing effects that reduce fine detail preservation. Conversely, a larger

λ fails to enforce sufficient low-frequency consistency, resulting in illumination inconsistencies across viewpoints.Our analysis reveals that $\lambda = 0.2$ achieves the best overall performance.

VI. CONCLUSION

We propose a frequency-guided framework for low-light novel view synthesis that addresses illumination inconsistencies and sensor noise through three key components: a Frequency-aware Lightness Modulator for detail-preserving illumination enhancement, an Adaptive Denoising Mask for filtering unstable Gaussians and reducing storage, and a Multi-view Frequency Consistency constraint for cross-view coherence. Experiments on the Gaussian-DK dataset demonstrate state-of-the-art performance with storage reduction and real-time rendering. Beyond low-light scenes, our frequency-domain guidance offers a promising direction for robust 3D reconstruction under other adverse conditions such as haze, rain, and motion blur.

REFERENCES

[1] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering." *ACM Trans. Graph.*, vol. 42, no. 4, pp. 139–1, 2023.

[2] S. Ye, Z.-H. Dong, Y. Hu, Y.-H. Wen, and Y.-J. Liu, "Gaussian in the dark: Real-time view synthesis from inconsistent dark images using gaussian splatting," in *Computer Graphics Forum*, vol. 43, no. 7. Wiley Online Library, 2024, p. e15213.

[3] R. Martin-Brualla, N. Radwan, M. S. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth, "Nerf in the wild: Neural radiance fields for unconstrained photo collections," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 7210–7219.

[4] X. Huang, Q. Zhang, Y. Feng, H. Li, X. Wang, and Q. Wang, "Hdr-nerf: High dynamic range neural radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18 398–18 408.

[5] M. Levoy and P. Hanrahan, "Light field rendering," in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 2023, pp. 441–452.

[6] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[7] H. Shum and S. B. Kang, "Review of image-based rendering techniques," *Visual Communications and Image Processing 2000*, vol. 4067, pp. 2–13, 2000.

[8] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[9] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-nerf 360: Unbounded anti-aliased neural radiance fields," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5470–5479.

[10] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM transactions on graphics (TOG)*, vol. 41, no. 4, pp. 1–15, 2022.

[11] A. Yu, V. Ye, M. Tancik, and A. Kanazawa, "pixelnerf: Neural radiance fields from one or few images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 4578–4587.

[12] J. Luiten, G. Kopanas, B. Leibe, and D. Ramanan, "Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis," in *2024 International Conference on 3D Vision (3DV)*. IEEE, 2024, pp. 800–809.

[13] Z. Yu, A. Chen, B. Huang, T. Sattler, and A. Geiger, "Mip-splatting: Alias-free 3d gaussian splatting," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 19 447–19 456.

[14] J. Lin, Z. Li, X. Tang, J. Liu, S. Liu, J. Liu, Y. Lu, X. Wu, S. Xu, Y. Yan, *et al.*, "Vastgaussian: Vast 3d gaussians for large scene reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5166–5175.

[15] Z. Fan, K. Wang, K. Wen, Z. Zhu, D. Xu, Z. Wang, *et al.*, "Lightgaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ fps," *Advances in neural information processing systems*, vol. 37, pp. 140 138–140 158, 2024.

[16] Y. Liu, Z. Zhong, Y. Zhan, S. Xu, and X. Sun, "Maskgaussian: Adaptive 3d gaussian representation from probabilistic masks," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 681–690.

[17] E. H. Land, "The retinex theory of color vision," *Scientific american*, vol. 237, no. 6, pp. 108–129, 1977.

[18] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.

[19] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 2023, pp. 643–652.

[20] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[21] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM international conference on multimedia*, 2019, pp. 1632–1640.

[22] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3291–3300.

[23] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE transactions on image processing*, vol. 30, pp. 2340–2349, 2021.

[24] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1780–1789.

[25] H. Zhou, W. Dong, and J. Chen, "Lita-gs: Illumination-agnostic novel view synthesis via reference-free 3d gaussian splatting and physical priors," 2025. [Online]. Available: https://arxiv.org/abs/2504.00219

[26] Z. Cui, X. Chu, and T. Harada, "Luminance-gs: Adapting 3d gaussian splatting to challenging lighting conditions with view-adaptive curve adjustment," in *CVPR*, 2025.

[27] H. Sun, F. Yu, H. Xu, T. Zhang, and C. Zou, "Ll-gaussian: Low-light scene reconstruction and enhancement via gaussian splatting for novel view synthesis," 2025. [Online]. Available: https://arxiv.org/abs/2504.10331

[28] A. Oppenheim and J. Lim, "The importance of phase in signals," in *IEEE Proceedings*, vol. 69, 1981, pp. 529–541.

[29] A. v. Oppenheim, R. Schafer, and T. Stockham, "Nonlinear filtering of multiplied and convolved signals," *IEEE transactions on audio and electroacoustics*, vol. 16, no. 3, pp. 437–466, 2003.

[30] T. G. Stockham, "Image processing in the context of a visual model," *Proc. IEEE*, vol. 60, no. 7, pp. 828–842, 1972.

[31] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[32] L. Jiang, B. Dai, W. Wu, and C. C. Loy, "Focal frequency loss for image reconstruction and synthesis," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 13 919–13 929.

[33] L. Chi, B. Jiang, and Y. Mu, "Fast fourier convolution," *Advances in Neural Information Processing Systems*, vol. 33, pp. 4479–4488, 2020.

[34] M. Xu, F. Zhan, J. Zhang, Y. Yu, X. Zhang, C. Theobalt, L. Shao, and S. Lu, "Wavenerf: Wavelet-based generalizable neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 18 195–18 204.

[35] J. Zhang, F. Zhan, M. Xu, S. Lu, and E. Xing, "Fregs: 3d gaussian splatting with progressive frequency regularization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 21 424–21 433.

[36] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.