LGDWT-GS: Local and Global Discrete Wavelet-Regularized 3D Gaussian Splatting for Sparse-View Scene Reconstruction

Shima Salehi, Atharva Agashe, Andrew J. McFarland, and Joshua Peeples, *Member, IEEE*

***Abstract*—We propose a new method for few-shot 3D reconstruction that integrates global and local frequency regularization to stabilize geometry and preserve fine details under sparse-view conditions, addressing a key limitation of existing 3D Gaussian Splatting (3DGS) models. We also introduce a new multispectral greenhouse dataset containing four spectral bands captured from diverse plant species under controlled conditions. Alongside the dataset, we release an open-source benchmarking package that defines standardized few-shot reconstruction protocols for evaluating 3DGS-based methods. Experiments on our multispectral dataset, as well as standard benchmarks, demonstrate that the proposed method achieves sharper, more stable, and spectrally consistent reconstructions than existing baselines. The dataset and code for this work are publicly available [1].**

***Index Terms*—3D Gaussian Splatting, Frequency-Domain Regularization, Discrete Wavelet Transform, Sparse-View Reconstruction, Multispectral 3D Imaging, Multispectral Reconstruction, Few-Shot Benchmarking Package**

I. Introduction

Novel view synthesis aims to render unseen viewpoints of a scene from a limited number of images. Neural Radiance Fields (NeRF) [1] achieve photorealistic quality but rely on dense multi-view supervision and long training times, which limits their use in practical settings such as robotics [2] and agriculture [3] domains where capturing many views is often infeasible. 3DGS [4] addresses these computational bottlenecks, enabling real-time rendering and efficient optimization. However, in few-view scenarios, 3DGS tends to over-reconstruct available high-frequency (HF) regions, sharply reproducing textures and edges in training views, while losing smooth low-frequency (LF) structures and overall stability of the scene [5].

We address the few-view 3D reconstruction challenge with our proposed method, a frequency-aware extension of 3DGS that integrates the Discrete Wavelet Transform (DWT) to guide spatial-frequency learning. By decomposing rendered and ground-truth images into multi-scale sub-bands, the model explicitly balances both lowand high frequency information. Two complementary supervision strategies are introduced: a global DWT loss that preserves large-scale consistency and a patch-wise DWT loss that refines local details and fine edges. Together, these components balance between HF and LF regions, yielding sharper, more reliable reconstructions under sparse supervision. Building on this foundation, we extend the framework to handle multispectral data. The proposed MultiSpectral 3DGS jointly reconstructs RGB and Near-Infrared (NIR) views using shared geometry and modality-specific appearance parameters, ensuring spectral and spatial coherence across bands. To support research in this area, we introduce a new multispectral greenhouse dataset along with an open-source few-shot benchmarking package that standardizes sparse-view evaluation protocols. We validate both the RGB and multispectral versions of LGDWT-GS on standard benchmarks (LLFF [6], MipNeRF360 [7]) and on our new dataset. The main contributions of this work are summarized as follows:

- Introduction of joint local and global frequency-domain supervision to improve 3D reconstruction.
- Development of an open-source multispectral greenhouse dataset containing four spectral bands (Red, Green, Red-Edge, NIR).
- Extension of 3DGS to the multispectral domain, enabling consistent cross-spectral reconstruction.
- Establishment of a standardized few-shot 3DGS benchmark and evaluation protocol.

II. Related Work

A. Few-Shot Novel View Synthesis

Building upon the foundational NeRF framework, recent methodologies have emerged to address the challenges of few-shot novel view synthesis. Early approaches in this domain typically rely on strong regularization priors. For instance, RegNeRF [8] regularizes geometry by enforcing smoothness constraints on unobserved viewpoints, while FreeNeRF [9] employs frequency regularization to prevent HF artifacts during the early stages of training. Other methods, such as SparseNeRF [10] and DietNeRF [11], incorporate auxiliary supervision to guide reconstruction. Specifically, SparseNeRF utilizes depth priors, while DietNeRF leverages semantic consistency losses to guide geometry in occluded regions.

Recent advancements have focused on robust geometric adaptation without heavy external priors. FrugalNeRF [12]

S. Salehi is a Ph.D. student, A. Agashe is a Master's student, and J. Peeples is an Assistant Professor in the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77843, USA e-mails: shima.salehi@tamu.edu, atharvagashe22@tamu.edu and jpeeples@tamu.edu

A. J. McFarland is a graduate student with the Department of Horticultural Sciences and the Texas A&M AgriLife Research Automated Precision Phenotyping Facility, Texas A&M University, College Station, TX, 77843, USA, e-mail: andrew.mcfarland@ag.tamu.edu

[1] https://github.com/Advanced-Vision-and-Learning-Lab/sparse-view-3dgs-pack

introduces a cross-scale sharing scheme to maximize information utility from limited pixels. Furthermore, addressing the practical reality of agricultural and robotic data, methods like SPARF [13] have extended few-shot capabilities to handle noisy camera poses, jointly refining extrinsic parameters and scene geometry to prevent drift in uncontrolled environments.

B. 3DGS and Sparse-View Extensions

3DGS substantially reduces the training latency of NeRF-based methods through efficient differentiable rasterization. Despite this advantage, 3DGS exhibits a characteristic failure mode in sparse-view or few-shot regimes. In such settings, the optimization process tends to over-reconstruct HF details, particularly edges and fine textures, in the observed training views. Due to limited multiview constraints, this behavior leads to poor generalization across viewpoints and results in degraded global geometric coherence and weakened structural consistency in unobserved regions [5].

To address these limitations, several extensions have been proposed to improve the robustness of 3DGS under sparse supervision. Methods such as FSGS [14] and PGDGS [15] employ adaptive and progressive densification strategies that incrementally populate underconstrained regions of the scene, thereby reducing geometric sparsity and improving reconstruction stability. Moreover, DNGaussian [16] and SCGaussian [17] focus on regularizing scene geometry through explicit depth constraints and structural consistency priors. By suppressing spurious or unstable Gaussian primitives, these approaches enhance geometric plausibility and reduce reconstruction artifacts in sparsely observed areas. While effective, such methods primarily rely on geometric supervision and do not explicitly regulate the spectral distribution of reconstructed content, leaving frequency-domain inconsistencies insufficiently constrained.

C. Frequency-Aware Supervision

Frequency-domain analysis has been adopted in neural rendering as an effective means to separate global structural information from fine-scale texture. LF components primarily encode smooth geometry and large-scale scene structure, whereas HF components correspond to edges and detailed appearance variations. By explicitly regulating these components, frequency-aware supervision provides a principled mechanism for stabilizing optimization under limited viewpoint coverage.WaveNeRF [18] and DWT-NeRF [5] use wavelet guidance to improve NeRF training. This reduces errors and sharpens edges, especially when there are few camera views. Similarly, DWT-GS [19] applies this concept to Gaussian Splatting to remove HF noise. However, these methods generally apply rules to the whole image at once. They fail to distinguish between preserving the main structure and refining fine details."

Building on these observations, our approach introduces a dual-branch frequency-aware supervision strategy that is directly integrated into the 3DGS rasterization pipeline. A global frequency branch enforces LF consistency to preserve overall geometric structure, while a complementary patch-wise branch selectively refines HF components to recover local details without inducing overfitting.

D. Multispectral and Agricultural 3D Reconstruction

Multispectral and hyperspectral imaging capture reflectance information beyond the visible spectrum, enabling critical applications in agriculture [20], remote sensing [21], [22], and material analysis [23]. By providing wavelength-dependent measurements, these modalities support robust characterization of vegetation health, structural properties, and material composition that cannot be inferred from RGB imagery alone. Several NeRF-based extensions have been proposed to model spectral radiance fields across multiple wavelengths. Methods such as HS-NeRF [24], SpectralNeRF [25], and Spec-NeRF [26] explicitly reconstruct spectral reflectance by conditioning radiance fields on wavelength information. More recently, HyperGS [27] adapts Gaussian Splatting to hyperspectral scenes, demonstrating the feasibility of point-based spectral rendering. However, these approaches generally assume dense multi-view acquisition and rely on high-cost sensing hardware, limiting their applicability in real-world agricultural settings.

In practice, agricultural imaging is often performed under few-view, spectrally heterogeneous conditions, with additional challenges arising from varying illumination, plant self-occlusions, and complex scene geometry. To better reflect these constraints, we introduce a controlled multispectral greenhouse dataset capturing Red, Green, Red Edge, and NIR channels, along with a unified few-shot benchmarking package. This dataset enables systematic evaluation of frequency-aware multispectral 3D reconstruction methods under realistic agricultural imaging conditions.

III. Method

We introduce the LGDWT-GS framework, which integrates frequency-domain supervision into the 3DGS pipeline to improve few-shot 3D reconstruction. Then, we extend this framework to a multispectral version that jointly reconstructs RGB and NIR modalities under a shared 3D geometry.

A. LGDWT-GS Framework

The proposed method extends the original 3DGS by introducing frequency-aware supervision through global and local (patch-wise) DWT losses. This integration enhances large-scale structural consistency and fine-grained texture recovery without modifying the differentiable splatting renderer. In few-shot settings, only a small number of sparse points can be reconstructed by COLMAP [28] due to limited view overlap. To obtain a denser and more stable initialization, depth priors and multi-view stereo reconstructions are used to generate additional pseudo-points, forming a more complete geometric basis for the Gaussian representation. Each Gaussian primitive is parameterized by a mean vector, covariance matrix, color, and opacity.

During training, the differentiable rasterizer projects these Gaussians into the image plane, and the rendered outputs are

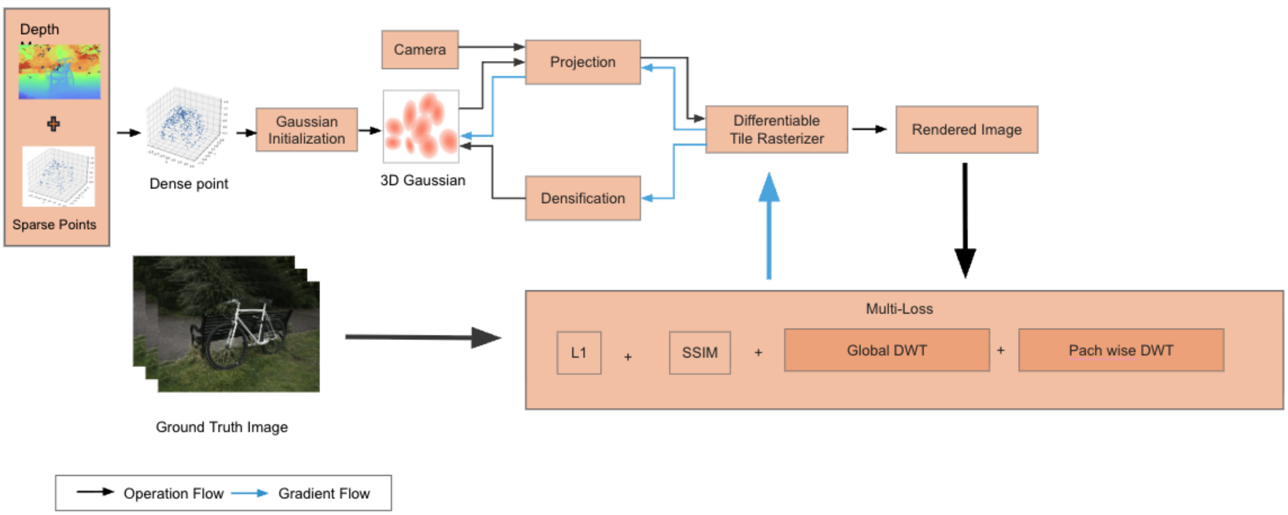

Fig. 1. Overview of the LGDWT-GS framework. The model introduces frequency-domain regularization through global and local DWT losses. Combined supervision (L1, SSIM, and global DWT and local DWT) enhances structural stability and textural fidelity. Black and blue arrows represent operation and gradient flows, respectively.

optimized against ground-truth views using a composite loss function,shown in Equation 1, that jointly balances spatial, perceptual, and frequency domain objectives

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{L1}} + \mathcal{L}_{\text{SSIM}} + \alpha\mathcal{L}_{\text{gDWT}} + \beta\mathcal{L}_{\text{pDWT}}. \tag{1}$$

$\mathcal{L}_{\text{L1}}$ ensures pixel-wise fidelity, $\mathcal{L}_{\text{SSIM}}$ enforces structural similarity, $\mathcal{L}_{\text{gDWT}}$ and $\mathcal{L}_{\text{pDWT}}$ regularizes global and local frequency components, respectively. The weighting factors α and β control the relative contribution of global and patch-wise DWT supervision during optimization. The overall proposed method with this loss function is shown in Figure 1.

1) Global DWT Supervision: The global DWT loss enforces large-scale frequency consistency between rendered and ground-truth images, promoting structural stability across views. Each image is decomposed using a one-level Haar wavelet transform into four frequency subbands. The LL subband represents the LF approximation of the image and preserves its coarse global structure. The LH subband captures horizontal HF components corresponding to vertical edges. The HL subband captures vertical HF components corresponding to horizontal edges. The HH subband contains diagonal HF information that highlights fine, corner-like details.

I_{LL} captures coarse structural information and global illumination, I_{LH} and I_{HL} represent horizontal and vertical edge responses, and I_{HH} encodes fine textures and noise. Fig. 2. The global DWT loss Equation 2 term is defined as follows:

$$\mathcal{L}_{\text{Global-DWT}} = \sum_{s \in \{LL,LH,HL,HH\}} w_s \, \|\widehat{\mathbf{I}}_S - \mathbf{I}_S\|_1, \tag{2}$$

where w_s denotes the frequency weight for each subband. To mitigate overfitting to unstable HF regions [9], the HH component is included but assigned a weight near zero ($w_{HH} \approx 0$). This formulation aligns global frequency structures between rendered and reference images, ensuring coherent reconstruction of both LF and mid-frequency content while suppressing noisy HF artifacts.

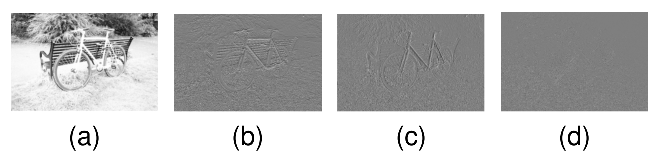

Fig. 2. Wavelet decomposition of the input image into four subbands: (a) LL (approximation), (b) LH (horizontal), (c) HL (vertical), (d) HH (diagonal).

2) Patch-wise DWT Supervision: Although global supervision enforces overall frequency consistency, local regions may still exhibit under-reconstruction, particularly where fine HF details are embedded within smooth, LF areas. To address this, patch-wise DWT supervision is introduced to refine these regions by locally emphasizing frequency balance. Rendered and ground-truth images are divided into non-overlapping patches of fixed size and stride. Each patch is independently analyzed to detect local frequency imbalance, guided by a Low-frequncy energy(E_{LF}) metric that quantifies the dominance of LF versus HF content Equation 3. Each image is first decomposed via a one-level Haar wavelet transform: where I_{LL} denotes the low frequency (structural) subband and I_{HF} represents the aggregated high frequency components (I_{LH}, I_{HL}, I_{HH}). The E_{LF} for each spatial location is defined as:

$$E_{LF}(x,y) = \frac{\|\mathbf{I}_{LL}(x,y)\|_1}{\|\mathbf{I}_{LL}(x,y)\|_1 + \|\mathbf{I}_{HF}(x,y)\|_1}, \tag{3}$$

Regions with low E_{LF} values, defined as pixels whose E_{LF} falls below a percentile based threshold that is treated as a hyperparameter and empirically set to the lowest 20% of the E_{LF} distribution per image, correspond to areas where

(a) (b)

Fig. 3. E_{LF} map used for patch selection: (a) Ground Truth, (b) E_{LF} Map. Red regions denote low E_{LF} values, indicating weak LF stability or missing HF details and revealing spatial frequency imbalance in the reconstruction.

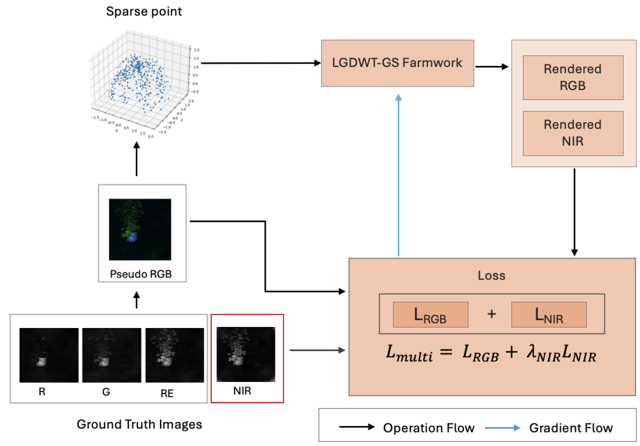

Fig. 4. Multispectral LGDWT-3DGS framework. Pseudo-RGB images (constructed from Red, Green, and Red-Edge bands) are used for COLMAP-based pose estimation and sparse reconstruction. RGB and NIR chanells are then jointly optimized under a shared geometry using cross-spectral supervision and DWT-based frequency regularization, improving geometric consistency and spectral alignment under sparse-view scenarios.

structural information is weak or where HF details are missing in LF regions, as illustrated in Figure 3. These regions are therefore prioritized for localized frequency refinement.

For each selected patch, one-level Haar wavelet transform is applied, and weighted losses are computed on the HF subbands to enhance local detail reconstruction. The patch-level loss Equation 4 is defined as:

$$\mathcal{L}_{\text{Patch-DWT}} = \frac{1}{N_p} \sum_{p=1}^{N_p} \sum_{B \in \{LH, HL\}} \|\widehat{\mathbf{I}}_B^p - \mathbf{I}_B^p\|_1, \tag{4}$$

where N_p denotes the number of selected patches and b the corresponding frequency band. This localized supervision improves fine-grained reconstruction by reinforcing HF corrections within LF regions, sharpening object boundaries, and restoring texture details while maintaining global stability.

B. MultiSpectral Extension

The LGDWT-GS framework is extended to support dual-modality reconstruction, jointly modeling RGB and NIR images under a shared 3D geometry. This multispectral formulation introduces cross-spectral supervision, improving both geometric fidelity and spectral alignment across modalities. To initialize the geometry, pseudo-RGB images are first generated by combining the Red, Green, and Red-Edge spectral channels. These pseudo-RGB images are used as inputs to COLMAP to estimate camera poses and generate the initial sparse point cloud. The resulting structure provides reliable geometric alignment across all spectral bands, even in sparse-view conditions.

The RGB and NIR data are stored in separate but spatially aligned folders, sharing identical intrinsic and extrinsic camera parameters. This setup ensures pixel-level correspondence between modalities and enables synchronized multispectral loading during training. Optional depth priors can also be incorporated to further densify the initialization and stabilize reconstruction in regions with limited multiview overlap. Each Gaussian primitive maintains shared geometric parameters while encoding modality-specific color attributes. Two-pass differentiable rasterization is then applied to produce the corresponding renderings. allowing both branches to be optimized jointly under a unified geometry while preserving spectral distinctions.

Supervision combines reconstruction objectives from both the RGB channel and the NIR channel Equation 5:

$$\mathcal{L}_{\text{Multi}} = \mathcal{L}_{\text{RGB}} + \lambda_{\text{NIR}}\, \mathcal{L}_{\text{NIR}}, \tag{5}$$

where it λ_{NIR} controls the relative contribution of the NIR branch. During densification, new Gaussians are spawned in regions exhibiting high residuals in either spectral channels Equation 6:

$$\text{mask} = \max(\text{RGB}_{\text{res}}, \text{NIR}_{\text{res}}), \tag{6}$$

ensuring that both spectral domains selectively guide geometric refinement. This shared-geometry, dual-appearance design promotes cross-spectral coherence and enhances fine-detail reconstruction under sparse-view conditions.

C. Dataset

To construct the proposed multispectral greenhouse dataset, we employed the MSIS-AGRI-1-A system: a snapshot multispectral camera equipped with four spectral bands centered at 580 nm (Green), 660 nm (Red), 735 nm (Red Edge), and 820 nm (NIR) Figure 5. The camera uses a global shutter CMOS sensor with 4 MP resolution and integrates Anti-X-Talk™ technology to minimize spectral leakage between channels, ensuring radiometrically accurate and high-contrast measurements. Each capture was synchronized with a four-channel LED illumination module operating at the same wavelengths to maintain consistent spectral lighting across all bands. Before data collection, both cameras were calibrated for intrinsic and extrinsic parameters using a checkerboard target to ensure precise geometric alignment.

Each scene corresponds to an individual plant species sorghum, tomato, alocasia, cotton, and grape captured inside a controlled greenhouse imaging station Figure 6. The setup includes a motorized turntable, a uniform black background, and a multispectral LED illumination system designed to ensure radiometric consistency and suppress ambient reflections.

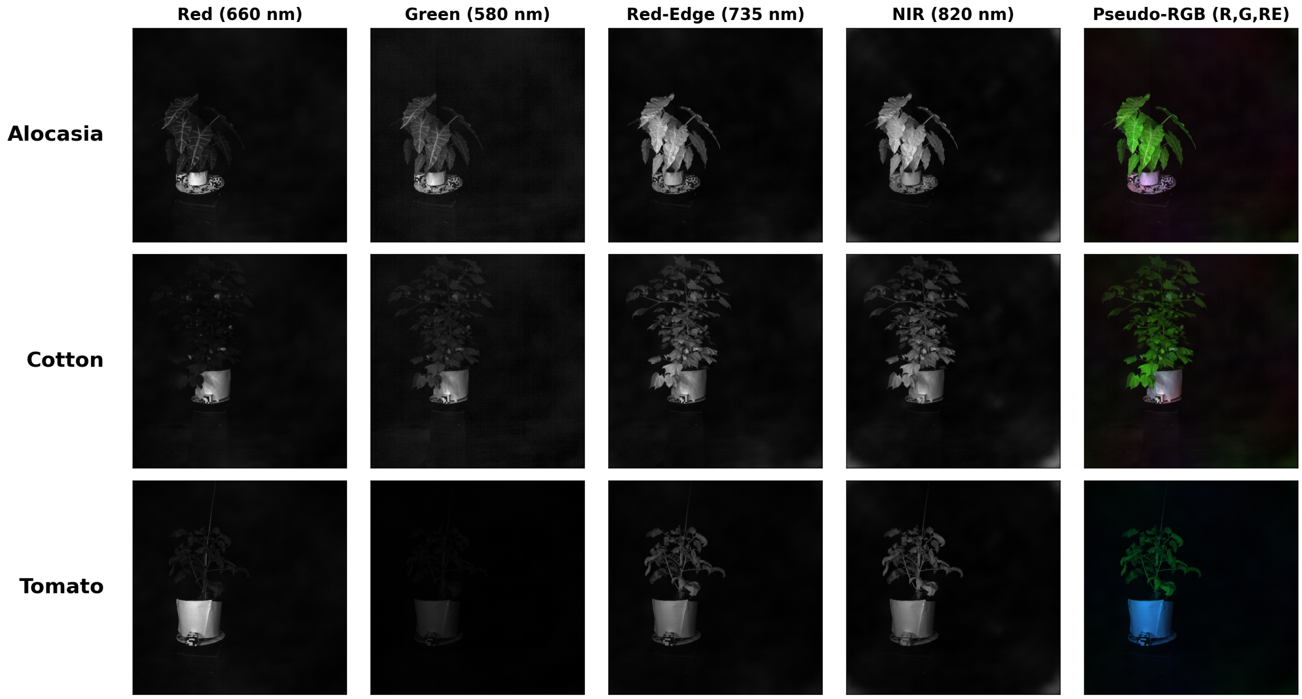

Fig. 5. Example spectral channels for three representative plant scenes. Columns correspond to 580 nm (Green), 660 nm (Red), 735 nm (Red Edge), and 820 nm (NIR) bands. The final column shows the pseudo-RGB composite used for COLMAP reconstruction.

Fig. 6. Greenhouse imaging setup. The MSIS-AGRI-1-A camera, LED illumination system, motorized turntable, and cube reference markers used for geometric calibration are shown.

Two identical cameras were placed on opposite sides of the turntable, providing dual viewpoints. For each fixed horizontal position, the cameras were vertically translated across four discrete heights to capture multiple canopy layers. Each plant was imaged at ten rotational steps (36° increments) using the motorized turntable, with two reference cube markers attached for rotation tracking and geometric calibration. Each acquisition session generated approximately 80–100 multispectral frames per scene, yielding almost 500 spatially aligned spectral images across all plant species.

For each acquisition, four spectral bands were recorded simultaneously, forming a co-registered multispectral image cube $\{I_R, I_G, I_{RE}, I_{NIR}\}$. A pseudo-RGB composite was generated from the Red, Green, and Red Edge channels to facilitate Structure-from-Motion (SfM) reconstruction using COLMAP. Reconstructed camera poses, sparse point clouds, and scene bounds were used as geometric initialization for multispectral 3DGS training. This preprocessing ensures spatial alignment between spectral modalities and enables consistent cross-spectral supervision during reconstruction.

IV. Experiments

A. Benchmarking

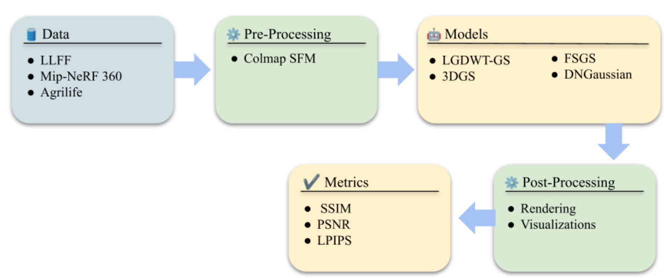

Fig. 7. Few-shot 3DGS benchmarking tool overview. The overall framework and figure is adapted from Anomalib [29].

We introduce an end-to-end benchmarking pipeline for few-shot 3D reconstruction that unifies data processing, training, and evaluation across multiple Gaussian Splatting baselines. While existing frameworks such as Nerfstudio [30] simplify NeRF-style experimentation, they do not yet provide native support for few-shot Gaussian Splatting workflows or standardized evaluation under sparse-view conditions. Our system uses common multi-view datasets and executes a unified SfM stage using COLMAP to ensure consistent camera poses and sparse geometry across all experiments. On top of this shared foundation, the pipeline supports the training of multiple few-shot Gaussian Splatting variants, including 3DGS [4], FSGS [14], and DNGaussian [16], using a common configuration and logging interface.

The benchmarking package is fully modular, allowing new datasets and new Gaussian Splatting methods to be registered through a unified API without re-engineering the environment or modifying existing pipelines. This design enables a one-install, many-model workflow with a fixed data layout, standardized metrics, and consistent evaluation protocols. As a result, comparisons across models, scenes, and random seeds become fair, repeatable, and reproducible, minimizing environment-dependent variability and facilitating reliable analysis of few-shot 3D reconstruction performance.

B. Quantitative Results

We evaluate the proposed framework on standard benchmarks including LLFF [6], MipNeRF360 [7], and our controlled greenhouse multispectral dataset. The evaluation is designed to assess reconstruction quality under varying levels of view sparsity and spectral diversity. All models are trained on a single NVIDIA A100 GPU. Despite its efficiency, the proposed LGDWT-GS pipeline converges in under three minutes per scene while achieving state-of-the-art performance across sparse-view, dense-view, and multispectral settings.

Although the proposed framework supports arbitrary numbers of input views, we report results under representative configurations that are commonly adopted in prior work. Specifically, we evaluate on LLFF using three input views, which is a standard few-shot setting. For MipNeRF360, we report results using 24 input views, which is widely treated as a few-shot configuration due to the dataset's complex geometry and wide viewpoint variation. For the greenhouse dataset, we adopt a fixed few-shot setting with ten views per plant, reflecting realistic agricultural capture conditions.

Table I shows that LGDWT-GS improves performance significantly on the three-view LLFF dataset. These results confirm that frequency supervision prevents the overfitting and structural errors often seen in sparse-view training. LGDWT-GS also consistently outperforms DNGaussian, suggesting that frequency control offers benefits that geometry modeling alone cannot provide. Additionally, our method achieves higher scores than NeRF-based methods like RegNeRF and FreeNeRF. This proves that enforcing consistency in the frequency domain preserves both global structure and fine details, even with very few views.

Table II presents results on MipNeRF360 using 24 input views. Even with more images, this dataset remains difficult due to complex scenes and wide angles. Under these conditions, LGDWT-GS achieves the highest overall scores. The strong performance against DNGaussian shows that frequency supervision is effective beyond just sparse data. It ensures stable reconstruction across both limited-view and complex multi-view scenarios.

TABLE I
COMPARISON ON LLFF (3-VIEW). METRICS INCLUDE PSNR (DB) ↑, SSIM ↑, AND LPIPS ↓.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Mip-NeRF360	15.22	0.351	0.540
DietNeRF	13.86	0.305	0.578
RegNeRF	18.66	0.535	0.411
FreeNeRF	19.13	0.562	0.384
SparseNeRF	19.07	0.564	0.392
3DGS	16.94	0.488	0.402
DNGaussian	19.73	0.669	0.301
LGDWT-GS (ours)	**20.46**	**0.726**	**0.279**

TABLE II
COMPARISON ON MIPNERF-360 (24-VIEW). METRICS INCLUDE PSNR (DB) ↑, SSIM ↑, AND LPIPS ↓.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Mip-NeRF360	19.78	0.530	0.431
DietNeRF	19.11	0.482	0.452
RegNeRF	20.55	0.546	0.398
FreeNeRF	21.04	0.587	0.377
SparseNeRF	21.13	0.600	0.389
3DGS	19.93	0.588	0.401
DNGaussian	22.13	0.676	0.301
LGDWT-GS (ours)	**22.41**	**0.692**	**0.298**

To validate LGDWT-GS in a realistic agricultural setting, we evaluated it on our custom multispectral greenhouse dataset. Each plant was captured from ten viewpoints using synchronized RGB and NIR cameras. Table III compares the performance against single-channel and standard multispectral baselines. Notably, unlike the LLFF and MipNeRF 360 experiments where high-frequency subbands were excluded, here we explicitly utilized high-frequency regularization. Since the background is irrelevant for this application, we tuned the parameters to focus on the HF components of the plant structure. This targeted approach allows the model to recover fine details, such as leaf veins, without being constrained by background noise. Consequently, LGDWT-GS consistently achieves the highest reconstruction quality across all scenes.

C. Qualitative Analysis

Figure 8 presents qualitative comparisons between baseline 3DGS and our LGDWT-GS on standard benchmarks. Frequency-aware supervision improves detail preservation, notably in foliage, edges, and thin structures. Under sparse views, standard 3DGS often exhibits texture blurring or missing details, while LGDWT-GS retains spectral and spatial coherence. This qualitative trend is consistent across LLFF and MipNeRF360 datasets.

Figure 9 presents a qualitative comparison on the multispectral greenhouse dataset to analyze the respective contributions of spectral diversity and frequency-domain supervision. We evaluate four reconstruction settings. The single-channel baseline exhibits over-smoothed textures and distorted edges due to limited spectral information. Incorporating DWT improves edge sharpness but does not fully resolve geometric inconsistencies. In contrast, multispectral training enhances global structural stability, though fine details remain blurred. By combining both components, LGDWT-GS achieves the best overall

TABLE III
QUANTITATIVE RESULTS ON THE MULTISPECTRAL GREENHOUSE DATASET.

Scene	Single			Single + DWT			Multispectral			Multispectral + DWT		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Cotton	28.74	0.820	0.422	28.96	0.829	0.424	30.55	0.873	**0.256**	**30.68**	0.874	0.258
Grape	29.93	0.888	0.361	29.28	0.843	0.362	30.62	**0.926**	0.176	**31.01**	0.925	**0.175**
Sorghum	28.06	0.738	0.555	28.24	0.741	0.555	30.57	0.890	0.402	**31.08**	0.894	**0.399**
Tomato	26.65	0.788	0.357	27.08	0.802	0.351	29.33	0.885	**0.212**	**29.59**	0.892	0.213
Houseplant	29.36	0.831	0.417	28.43	0.799	0.421	29.24	0.875	**0.245**	**29.43**	0.875	0.247
Average	28.55	0.813	0.422	28.39	0.802	0.422	30.06	0.890	0.258	**30.51**	**0.892**	**0.258**

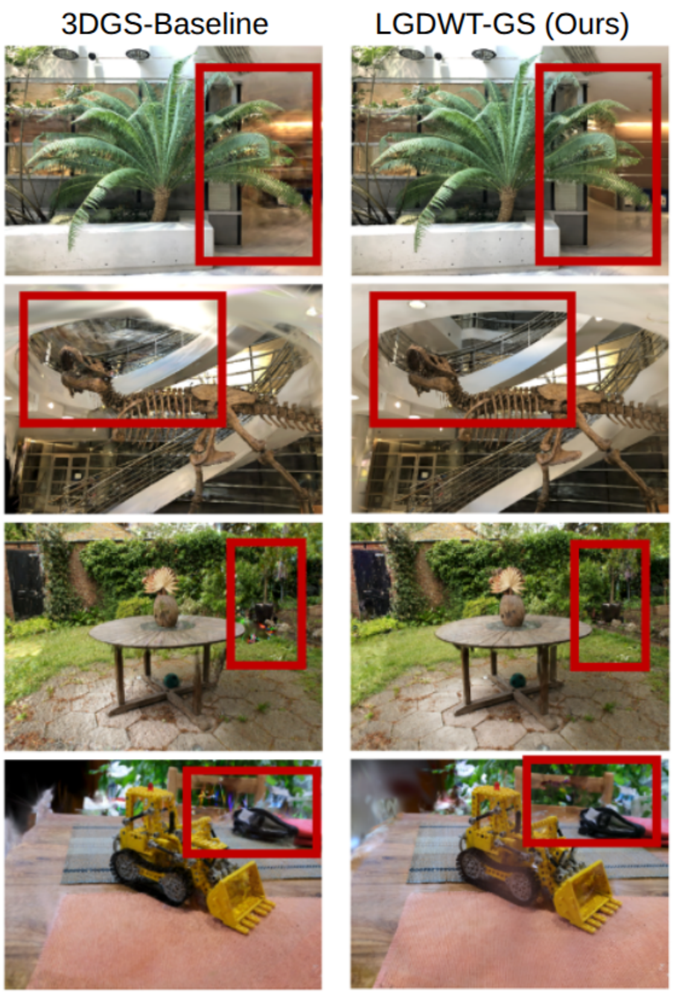

Fig. 8. Qualitative comparison between baseline 3DGS and our LGDWT-GS on LLFF and MipNeRF360. Red boxes highlight regions with better preservation of fine details (foliage, edges, thin structures).

performance, producing sharp and coherent reconstructions with well-preserved leaf boundaries and minimal artifacts. These results demonstrate that spectral diversity promotes geometric stability, while frequency supervision effectively recovers HF details.

D. Ablation Study

To isolate the contribution of frequency-domain supervision, we conducted a systematic component analysis on the 3-view LLFF configuration (Table IV). We compared our method against Neural Edge Histogram Descriptors (NEHD) [31]. NEHD uses Sobel kernels [32] to extract edge responses and aggregates the responses using differentiable histograms of gradient magnitudes to align the edge distributions of rendered and ground-truth images to improve texture representations. During optimization, NEHD explicitly focuses the rendering loss on HF edge regions. By heavily weighting these fine detail boundaries, NEHD forces the model to prioritize sharp discontinuities, achieving a PSNR of 19.99 dB. However, this aggressive focus on edges creates a trade-off: while it sharpens high-contrast boundaries, it often treats subtle surface textures like leaf veins as noise or fails to register them if they lack strong gradient magnitude. This results in reconstructions with sharp outlines but "waxy", over-smoothed interiors.

TABLE IV
ABLATION STUDY ON LLFF (3-VIEW) SCENES

Configuration	PSNR (dB)↑	SSIM↑	LPIPS↓	Time (s)
NEHD Loss	19.99	0.755	0.268	∼360
DWT Loss	19.92	0.680	0.314	∼120
DWT + Depth Reg.	20.03	0.683	0.304	∼126
DWT Staging	20.08	0.720	0.298	∼120
Two-Level DWT	20.28	0.726	0.297	∼166
Global + Local DWT	20.46	0.726	0.279	∼166

In contrast to NEHD, LGDWT-GS employs a DWT loss to supervise the entire frequency spectrum. While NEHD treats non-edge pixels uniformly, the DWT decomposes the signal into distinct sub-bands, effectively decoupling global structure from fine texture. This allows the model to enforce structural consistency through LF bands while simultaneously recovering detailed textures via HF bands even in regions with weak spatial gradients. Furthermore, in sparse-view settings, the DWT acts as a soft regularizer, suppressing HF artifacts in empty space while preserving valid signals in textured areas.

The component analysis highlights the specific contributions of our architectural choices. Depth regularization improves PSNR by 0.11 dB by stabilizing geometry and preventing Gaussians from drifting near the camera, a common failure in sparse SfM initialization. Additionally, employing a two-level decomposition with progressive frequency staging enables the model to establish coarse geometry before refining fine details, thereby preventing early convergence to local minima. Finally, combining global structural constraints with patch-based local refinement yields the most balanced performance, achieving the highest SSIM of 0.726 and an LPIPS of 0.279.

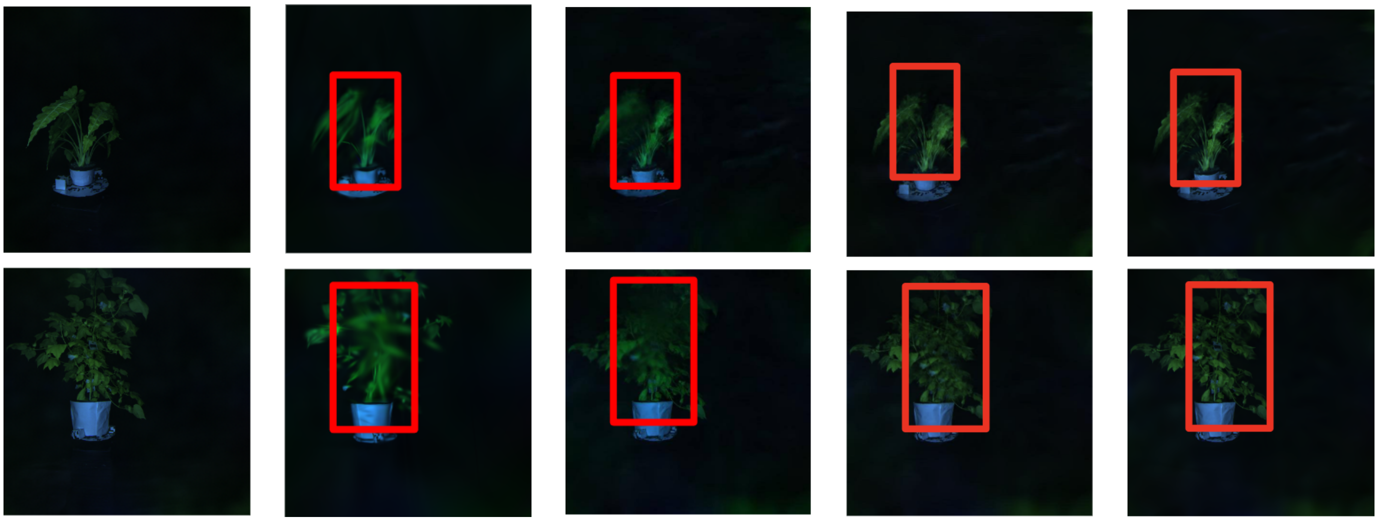

Fig. 9. Comparison of reconstruction outputs across input configurations: (a) Ground Truth, (b) Single-channel, (c) Single-channel + DWT, (d) Multispectral (no DWT), (e) Multispectral + DWT.

V. Conclusion

We introduced LGDWT-GS, a frequency-aware extension of 3DGS designed for few-shot 3D reconstruction. By integrating global and patch-wise DWT supervision, our method captures multiscale structural and textural information, mitigating the overreconstruction tendencies of sparse-view setups. The model preserves both large-scale consistency and fine-grained detail, achieving superior stability and perceptual quality compared to strong baselines such as 3DGS, SparseNeRF, and PGDGS across LLFF, MipNeRF360, and our new greenhouse datasets. Beyond the RGB domain, we extended the framework to a multispectral setting that jointly reconstructs RGB and NIR channels under shared geometry, ensuring spectral and spatial coherence across bands. To support research in this direction, we released a controlled multispectral greenhouse dataset and an accompanying few-shot benchmarking package that standardizes sparse-view evaluation protocols.

Overall, LGDWT-GS demonstrates that incorporating frequency-domain priors and multispectral supervision is an effective strategy for constructing efficient, detail-preserving 3D scene representations under data-limited conditions. Future work will focus on extending this framework with frequency-guided densification and pruning strategies to adaptively refine Gaussian distributions based on spectral frequency cues and to generalize the approach to broader multispectral datasets. Additionally, we aim to extend this framework to "in-the-wild" or less controlled agricultural environments, such as field-grown crops, to enable robust real-world 3D reconstruction under challenging outdoor conditions [33].

Acknowledgment

This material is based upon work supported by the Texas A&M University System Nuclear Security Office. Portions of this research were conducted with the advanced computing resources provided by Texas A&M High Performance Research Computing.

References

[1] 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.

[2] M. Z. Irshad, M. Comi, Y.-C. Lin, N. Heppert, A. Valada, R. Ambrus, Z. Kira, and J. Tremblay, "Neural fields in robotics: A survey," *arXiv preprint arXiv:2209.04310*, 2022.

[3] Y. Chen, B. Wang, Y. Wu, M. Zhao, T. Li, and Z. Zhang, "High-fidelity 3d reconstruction of plants using neural radiance fields," in *2022 IEEE International Conference on Robotics and Automation (ICRA)*, 2022, pp. 12 830–12 836.

[4] 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.

[5] H. Nguyen, R. Li, and T. Nguyen, "Dwtnerf: Boosting few-shot neural radiance fields via discrete wavelet transform," *arXiv preprint arXiv:2501.12637*, 2025.

[6] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. K. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–14, 2019.

[7] J. T. Barron, B. Mildenhall, M. Tancik, P. P. Srinivasan, R. Ramamoorthi, and R. Ng, "Mip-nerf 360: Unbounded anti-aliased neural radiance fields," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[8] M. Niemeyer, J. T. Barron, B. Mildenhall, A. Geiger, M. S. Sajjadi, and V. Larsson, "Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5480–5490.

[9] B. Yang, S. Peng, Y. Xu, Y. Shen, H. Bao, and X. Zhou, "Freenerf: Improving few-shot neural rendering with free frequency regularization," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 8254–8263.

[10] S. Roessle, P. P. Srinivasan, J. T. Barron *et al.*, "Sparsenerf: Distilling depth priors for efficient sparse-view novel view synthesis," *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[11] A. Jain, M. Tancik, and P. Abbeel, "Putting nerf on a diet: Semantically consistent few-shot view synthesis," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 5885–5894.

[12] C.-Y. Lin, C.-H. Wu, C.-H. Yeh, S.-H. Yen, C. Sun, and Y.-L. Liu, "Frugalnerf: Fast convergence for extreme few-shot novel view synthesis without learned priors," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 11 227–11 238.

[13] P. Truong, M.-J. Rakotosaona, F. Manhardt, and F. Tombari, "Sparf: Neural radiance fields from sparse and noisy poses," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4190–4200.

[14] Z. Zhu, Z. Fan, Y. Jiang, and Z. Wang, "Fsgs: Real-time few-shot view synthesis using gaussian splatting," in *European conference on computer vision*. Springer, 2024, pp. 145–163.

[15] H. Huang, Z. Zhang, G. Wu, and R. Wang, "Pgdgs: Improving few-shot 3d gaussian splatting with progressive gaussian densification," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[16] J. Li, J. Zhang, X. Bai, J. Zheng, X. Ning, J. Zhou, and L. Gu, "Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 20 775–20 785.

[17] R. Peng, W. Xu, L. Tang, L. Liao, J. Jiao, and R. Wang, "Structure consistent gaussian splatting with matching prior for few-shot novel view synthesis," *Advances in Neural Information Processing Systems*, vol. 37, pp. 97 328–97 352, 2024.

[18] 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.

[19] H. Nguyen, R. Li, A. Le, and T. Nguyen, "Dwtgs: Rethinking frequency regularization for sparse-view 3d gaussian splatting," 2025. [Online]. Available: https://arxiv.org/abs/2507.15690

[20] A. Karukayil, J. F. Mota, and F. A. Cheein, "3d crop reconstruction: A review of hyperspectral and multispectral approaches," *Computers and Electronics in Agriculture*, vol. 228, p. 109562, 2025.

[21] A. Ziemann and Z. Hampel-Arias, "New methods for new space: Multi-sensor change detection in remote sensing imagery," in *Pattern Recognition and Computer Vision in the New AI Era*. World Scientific, 2025, pp. 161–187.

[22] F. O. Nia, A. Mohammadi, S. A. Kharsa, P. Naikare, Z. Hampel-Arias, and J. Peeples, "Neighborhood feature pooling for remote sensing image classification," *arXiv preprint arXiv:2510.25077*, 2025.

[23] N. Klein, A. Carr, Z. Hampel-Arias, A. Ziemann, and E. Flynn, "Physics-guided neural networks for hyperspectral target identification," in *Applications of Machine Learning 2023*, vol. 12675. SPIE, 2023, p. 1267503.

[24] G. Chen, S. K. Narayanan, T. G. Ottou, B. Missaoui, H. Muriki, C. Pradalier, and Y. Chen, "Hyperspectral neural radiance fields," *arXiv preprint arXiv:2403.14839*, 2024.

[25] R. Li, J. Liu, G. Liu, S. Zhang, B. Zeng, and S. Liu, "Spectralnerf: Physically based spectral rendering with neural radiance field," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 4, 2024, pp. 3154–3162.

[26] J. Li, Y. Li, C. Sun, C. Wang, and J. Xiang, "Spec-nerf: Multi-spectral neural radiance fields," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 2485–2489.

[27] C. Thirgood, O. Mendez, E. Ling, J. Storey, and S. Hadfield, "Hypergs: Hyperspectral 3d gaussian splatting," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 5970–5979.

[28] J. L. Schönberger and J.-M. Frahm, "Structure–from–motion revisited," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[29] S. Akcay, D. Ameln, A. Vaidya, B. Lakshmanan, N. Ahuja, and U. Genc, "Anomalib: A deep learning library for anomaly detection," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 1706–1710.

[30] M. Tancik, E. Weber, E. Ng, R. Li, B. Yi, J. Kerr, T. Wang, A. Kristoffersen, J. Austin, K. Salahi, A. Ahuja, D. McAllister, and A. Kanazawa, "Nerfstudio: A modular framework for neural radiance field development," in *ACM SIGGRAPH 2023 Conference Proceedings*, 2023, pp. 1–12.

[31] J. Peeples, S. Al Kharsa, L. Saleh, and A. Zare, "Histogram layers for neural "engineered" features," *IEEE Transactions on Artificial Intelligence*, 2025.

[32] I. Sobel and G. Feldman, "A 3x3 isotropic gradient operator for image processing," *a talk at the Stanford Artificial Project in*, pp. 271–272, 1968.

[33] D. Zhang, J. Gajardo, T. Medic, I. Katircioglu, M. Boss, N. Kirchgessner, A. Walter, and L. Roth, "Wheat3dgs: In-field 3d reconstruction, instance segmentation and phenotyping of wheat heads with gaussian splatting," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 5360–5370.